



## On the consistency of bootstrap testing for a parameter on the boundary of the parameter space

Cavaliere, Giuseppe; Nielsen, Heino Bohn; Rahbek, Anders

*Published in:*  
Journal of Time Series Analysis

*DOI:*  
[10.1111/jtsa.12214](https://doi.org/10.1111/jtsa.12214)

*Publication date:*  
2017

*Document version*  
Peer reviewed version

*Document license:*  
[CC BY-NC-ND](#)

*Citation for published version (APA):*  
Cavaliere, G., Nielsen, H. B., & Rahbek, A. (2017). On the consistency of bootstrap testing for a parameter on the boundary of the parameter space. *Journal of Time Series Analysis*, 38(4), 513–534 .  
<https://doi.org/10.1111/jtsa.12214>

Giuseppe Cavaliere, Heino Bohn Nielsen,  
Anders Rahbek

On the Consistency of Bootstrap Testing  
for a Parameter on the Boundary of the  
Parameter Space

Quaderni di Dipartimento

Serie Ricerche 2016, n. 6

ISSN 1973-9346



ALMA MATER STUDIORUM  
UNIVERSITÀ DI BOLOGNA

Dipartimento di Scienze Statistiche “Paolo Fortunati”

# ON THE CONSISTENCY OF BOOTSTRAP TESTING FOR A PARAMETER ON THE BOUNDARY OF THE PARAMETER SPACE

GIUSEPPE CAVALIERE\*, HEINO BOHN NIELSEN<sup>†</sup>, AND ANDERS RAHBK<sup>‡</sup>

## ABSTRACT

It is well-known that with a parameter on the boundary of the parameter space, such as in the classic cases of testing for a zero location parameter or no ARCH effects, the classic nonparametric bootstrap – based on unrestricted parameter estimates – leads to inconsistent testing. In contrast, we show here that for the two aforementioned cases a nonparametric bootstrap test based on parameter estimates obtained under the null – referred to as ‘restricted bootstrap’ – is indeed consistent. While the restricted bootstrap is simple to implement in practice, novel theoretical arguments are required in order to establish consistency. In particular, since the bootstrap is analyzed both under the null hypothesis and under the alternative, non-standard asymptotic expansions are required to deal with parameters on the boundary. Detailed proofs of the asymptotic validity of the restricted bootstrap are given and, for the leading case of testing for no ARCH, a Monte Carlo study demonstrates that the bootstrap quasi-likelihood ratio statistic performs extremely well in terms of empirical size and power for even remarkably small samples, outperforming the standard and bootstrap Lagrange multiplier tests as well as the asymptotic quasi-likelihood ratio test.

KEYWORDS: Bootstrap; Boundary; ARCH; Location model.

JEL CLASSIFICATION: C32.

## 1 INTRODUCTION

In economics and econometrics, it is a well perceived fact that the finite sample properties of hypothesis tests, when based on asymptotic inference, can be quite poor. In such circumstances, it is also known that the bootstrap, when correctly implemented, can be an important device for improving upon the finite sample size properties of such tests. In addition, the bootstrap can also be an effectively tool for retrieving the asymptotic

---

\*Department of Statistical Sciences, University of Bologna, Italy. <sup>†</sup>Department of Economics, University of Copenhagen, Denmark. <sup>‡</sup>Department of Economics, University of Copenhagen, and CREATES, Denmark. We thank two referees as well as seminar and conference participants at University Paris Dauphine (MAF 2016), Oxford University, Toulouse School of Economics, University of Lecce and University of Salerno (SIS meeting 2016), for useful comments. This paper was written while Cavaliere was affiliated with the University of Copenhagen, Department of Economics, as visiting Professor. Funding from the Danish Council of Independent Research is gratefully acknowledged (Sapere Aude, Advanced Grant no. 12-124980). Correspondence to: Giuseppe Cavaliere, Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, 40125 Bologna, Italy; email: giuseppe.cavaliere@unibo.it.

distributions of tests statistics when such distributions depend on unknown parameters or do not have closed form expressions; see e.g. Davidson and MacKinnon (2006) and references therein.

Many important econometric testing problems with considerable relevance in applications, however, involve parameters which are on a boundary of a parameter space (as often defined by inequality, or mixed equality/inequality, constraints) under the null hypothesis. Examples include tests for no ARCH effects (Andrews, 2001) and tests for parameter constancy in random coefficient models (Andrews, 2001; Carrasco and Gregoir, 2002). When a parameter is on a boundary – such that standard regularity conditions fail to hold, leading in general to non-standard inference (Gouriéroux, Holly and Monfort, 1982) – it is well understood that the bootstrap is invalid, not only for obtaining asymptotic refinements but even for consistent estimation of the (first-order) asymptotic distributions of interest. See e.g. Horowitz (2001, p.3169), or Heckman, Smith and Clements (1997, p.530). A formal counterexample on the validity of the bootstrap is given in Andrews (2000) who shows, for a simple location model with one dimensional parameter space defined by an inequality constraint, invalidity of the classical nonparametric bootstrap – that is, the bootstrap based on *unrestricted* parameter estimates.

In this paper we contribute to this literature by focusing on *testing* the hypothesis that a parameter lies on the boundary of the parameter space. We establish that an alternative and simple nonparametric, ‘restricted bootstrap’ test, i.e. based on parameter estimates obtained with the null hypothesis imposed, can indeed be a useful and effective method for consistent testing on the boundary of the parameter space. We do so in terms of two well-known examples. The first is, as mentioned, the case of testing for  $\alpha = 0$ , where  $\alpha$  is a location parameter with bounded parameter space  $\alpha \geq 0$  for a sample of i.i.d. random variables with known scale; see Andrews (2000). The second deals with testing for a constant parameter in random coefficient models or, equivalently in terms of the Gaussian ARCH model, testing for no ARCH effects. This example is a special case of testing for no GARCH effects as discussed in Andrews (2001). For both cases we establish validity of the restricted bootstrap test, both under the null hypothesis and, importantly, also under the alternative. While this turns out to be straightforward for the location model, it is more involved for the ARCH case where validity of the our bootstrap test is achieved by verifying non-standard regularity conditions for inference on parameters on the boundary of the parameter space (Andrews, 1999, 2001), which are here modified for the application of our new bootstrap theory arguments. In particular, in the bootstrap case the non-standard asymptotic expansions required to deal with parameters on the boundary involve the introduction of pseudo-true parameters for the bootstrap data generating process. This is because in order to establish bootstrap validity, such non-standard asymptotic expansions have to be analyzed both under the null and crucially, under the alternative. While this is not strictly required for standard asymptotic inference, it is however a key step of the proof of validity of the bootstrap test. The use of the restricted bootstrap as a promising alternative approach to the classic bootstrap was studied for a different case of

non-standard inference in terms of nonstationarity and co-integration analysis in Cavaliere, Rahbek and Taylor (2012), Cavaliere, Nielsen and Rahbek (2015) and Boswijk, Cavaliere, Rahbek and Taylor (2015). Notably, the theory therein does not allow for inference with parameters on the boundary of the parameter space.

It is worth stressing that, with respect to Andrews (2000), our focus is not on bootstrap approximations of the distribution of an estimator when a parameter may or may not lie on the boundary of the parameter space. Instead, we are interested in bootstrap-based (likelihood ratio) testing of the hypothesis that a parameter lies *on* the boundary of the parameter space. While Andrews' (2000) results imply that the standard, unrestricted bootstrap cannot be used for testing an hypothesis of this kind, in contrast we show for the two cases previously mentioned that the *restricted* bootstrap can indeed be successfully used in order to construct a consistent test.

The paper is organized as follows. Section 2 establishes consistency of bootstrap hypothesis testing in the location model with restricted parameter space, while in Section 3 this is generalized to the case of no ARCH effects. In Section 4 Monte Carlo simulations provide empirical rejection frequencies under the null as well as under the alternative for the quasi-likelihood test of no ARCH effects. The study includes  $t$ -distributed as well as Gaussian innovations, and compare the proposed restricted bootstrap with the classical (standard and bootstrap) Lagrange multiplier test (see Engle, 1982) and the asymptotic test based on Andrews (2001). Section 5 concludes. All proofs and hence the new asymptotic bootstrap arguments needed for the results in Section 3 are collected in the appendix.

NOTATION: We use  $P^*$ ,  $E^*$  and  $V^*$  respectively to denote probability, expectation and variance, conditional on the original sample. With  $\xrightarrow{w} (\xrightarrow{w^*}_p)$  and  $\xrightarrow{p}$  we denote weak convergence (in probability), and convergence in probability, respectively, as the sample size  $T$  diverges. Moreover, for a given sequence  $X_T^*$  computed from the bootstrap data,  $X_T^* - X = o_p^*(1)$ , in probability, or  $X_T^* \xrightarrow{p^*}_p X$ , mean that for any  $\epsilon > 0$ ,  $P^*(\|X_T^* - X\| > \epsilon) \xrightarrow{p} 0$ , as  $T \rightarrow \infty$ . Similarly,  $X_T^* = O_p^*(1)$  in probability means that, for every  $\epsilon > 0$ , there exists a constant  $M > 0$  such that, for all large  $T$ ,  $P(P^*(\|X_T^*\| > M) < \epsilon)$  is arbitrarily close to one. Also  $\stackrel{d}{=}$  denotes equality in distribution. Finally,  $\mathbb{I}(\cdot)$  denotes the indicator function and  $x := y$  means that  $x$  is defined by  $y$ .

## 2 LOCATION MODEL

As the first motivating simple example consider the location model given by

$$X_t = \alpha + \varepsilon_t,$$

for  $t = 1, 2, \dots, T$ , and where  $\{\varepsilon_t : t \geq 1\}$  denotes a sequence of independent and identically distributed (i.i.d.)  $N(0, 1)$  random variables. The parameter space for  $\alpha$  is given by  $\Theta := \{\alpha \in \mathbb{R} \mid \alpha \geq 0\}$ , and the hypothesis of interest is  $H_0 : \alpha = 0$ . The unrestricted (Gaussian) maximum likelihood estimator (MLE) of  $\alpha$  is given by  $\hat{\alpha}_T = \max\{0, \bar{X}_T\}$ ,

where  $\bar{X}_T := T^{-1} \sum_{t=1}^T X_t$ , while the restricted MLE is  $\tilde{\alpha}_T = 0$ . The likelihood ratio (LR) statistic  $Q_T$  for  $H_0$  satisfies, under  $H_0$  and as  $T \rightarrow \infty$ ,

$$Q_T = T\hat{\alpha}_T^2 \xrightarrow{w} Q_\infty := (\max\{0, Z\})^2, \quad (1)$$

where  $Z$  is  $N(0, 1)$  distributed; see Andrews (2001). Note also that under the alternative,  $Q_T \rightarrow \infty$  at the rate of  $T$ .

The *unrestricted* (or classic) bootstrap addressed in Andrews (2000) generates the bootstrap process  $\{X_t^* : t \leq T\}$  as

$$X_t^* = \hat{\alpha}_T + \varepsilon_t^*, \quad (2)$$

where  $\varepsilon_t^*$  is sampled independently (conditional on the data) and with replacement from the (possibly centered) unrestricted residuals  $\{\hat{\varepsilon}_t : t \leq T\}$ ,  $\hat{\varepsilon}_t := X_t - \hat{\alpha}_T$ , and is based on the LR statistic of the simple null hypothesis that  $\alpha = \hat{\alpha}_T$ . The elegant arguments in Andrews (2000) imply that the bootstrap LR statistic does not converge (weakly in probability) to  $Q_\infty$  under  $H_0$ .

Consider now a *restricted* bootstrap test, based on estimation under  $H_0$ . In this case the bootstrap sample  $\{X_t^* : t \leq T\}$  is generated by

$$X_t^* = \tilde{\alpha}_T + \varepsilon_t^* = \varepsilon_t^*, \quad (3)$$

where  $\tilde{\alpha}_T = 0$  is the restricted MLE of  $\alpha$  and  $\varepsilon_t^*$  is sampled with replacement from the centered residuals imposing the null,  $\{\tilde{\varepsilon}_t - \tilde{\varepsilon}_T : t \leq T\}$  with  $\tilde{\varepsilon}_T := T^{-1} \sum_{t=1}^T \tilde{\varepsilon}_t$  and  $\tilde{\varepsilon}_t := X_t - \tilde{\alpha}_T = X_t$ . The bootstrap unrestricted MLE of  $\alpha$  is  $\hat{\alpha}_T^* = \max\{0, \bar{X}_T^*\}$ , while the restricted is  $\tilde{\alpha}_T^* = 0$  such that the bootstrap LR statistic for  $H_0$  can be explicitly expressed as

$$Q_T^* = T\hat{\alpha}_T^{*2}. \quad (4)$$

By Theorem 1 below,  $Q_T^*$  converges weakly in probability to  $Q_\infty$  for *any* value of the true parameter  $\alpha_0 \in \Theta$ . That is, the restricted bootstrap is consistent in the sense that the bootstrap mimics the correct (null) limit distribution  $Q_\infty$  under the null  $\alpha_0 = 0$ , while remaining bounded in probability under the alternative. Note that Theorem 1 does not require  $\{X_t : t \geq 1\}$  to be Gaussian, and the restricted bootstrap LR test can thus be seen as a valid quasi maximum likelihood (QML) test.

**THEOREM 1** *Let  $\{X_t : t \geq 1\}$  be an i.i.d. sequence with  $E(X_t) = \alpha_0 \in \Theta$  and  $V(X_t) = 1$ . Then for the bootstrap LR statistic defined in (4), as  $T \rightarrow \infty$ ,  $Q_T^* \xrightarrow{w^*}_p Q_\infty$  where  $Q_\infty$  is given by (1).*

The proof of Theorem 1 follows by applying standard results from bootstrap theory. Specifically, the result holds as by definition  $Q_T^* = T\hat{\alpha}_T^{*2} = (\max\{0, T^{1/2}\bar{X}_T^*\})^2$  with  $T^{1/2}\bar{X}_T^* \xrightarrow{w^*}_p Z$  using e.g. Theorem 23.4 in van der Vaart (1998).

**REMARK 2.1** Theorem 1 implies that the restricted bootstrap is consistent, since (i) under  $H_0$ , it replicates the correct null limiting distribution, and (ii) when  $H_0$  does not

hold,  $Q_T^* = O_p^*(1)$  (in probability) while  $Q_T$  diverges to infinity, thereby implying that the bootstrap test rejects  $H_0$  with probability tending to one. The result in Theorem 1 is even stronger, since it states that the bootstrap replicates the correct null asymptotic distribution even when the null is false.

REMARK 2.2 Theorem 1 also holds if the bootstrap shocks  $\varepsilon_t^*$  in (3) are obtained by i.i.d. resampling of the (centred) unrestricted residuals,  $\{\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}_T : t \leq T\}$  with  $\bar{\hat{\varepsilon}}_T := T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t$  and  $\hat{\varepsilon}_t := X_t - \hat{\alpha}_T$ , or if they are obtained parametrically (for instance, if  $\varepsilon_t^*$ , conditionally on the original data, is i.i.d.  $N(0, 1)$ , see also Andrews, 2000, p.402).

REMARK 2.3 It is worth noting that Andrews (2000) provides alternative methods for estimating the asymptotic distribution of the unrestricted estimator  $\hat{\alpha}_T$ . For instance, he shows that the ‘ $m$  out of  $n$ ’ bootstrap (where the bootstrap sample is  $m$ , with  $m = o(T)$ ) consistently estimates the distribution of  $\hat{\alpha}_T$  irrespectively of the true parameter to be on the boundary or not. A further alternative solution is to apply the unrestricted bootstrap to a standard, one-sided  $t$  test, which would be asymptotically valid for nominal levels below  $1/2$ . Unreported simulations show that when the emphasis is on testing  $H_0$ , the restricted bootstrap proposed here is the one performing better in terms of finite sample size and power.

### 3 ARCH

Next, we discuss the problem of testing for no ARCH effects. Although the analysis of validity of the bootstrap for the ARCH case requires new non-standard arguments and results (in particular in order to show validity of the bootstrap under the alternative), bootstrap consistency is established under mild assumptions.

Consider the ARCH model of order one for  $\{X_t : t \geq 1\}$  as given by

$$X_t = \sigma_t(\theta) z_t, \quad \sigma_t^2(\theta) = \omega + \alpha X_{t-1}^2, \quad \theta = (\alpha, \omega)' \quad (5)$$

with  $X_0$  fixed in the statistical analysis and where  $\{z_t : t \geq 1\}$  is an i.i.d. sequence with  $E(z_t) = 0$  and  $V(z_t) = 1$ . The hypothesis of interest is  $H_0 : \alpha = 0$ , which corresponds to the case of no ARCH effects.

Using the same framework as in Andrews (2001) for GARCH and random coefficient models, which nest our model as a special case, we let  $L_T(\theta)$  denote (up to a constant) the Gaussian log-likelihood function based on  $T$  observations,

$$L_T(\theta) := \sum_{t=1}^T \ell_t(\theta), \quad \ell_t(\theta) = -\frac{1}{2} \left( \log \sigma_t^2(\theta) + \frac{X_t^2}{\sigma_t^2(\theta)} \right). \quad (6)$$

The parameter space is defined as  $\Theta := \{(\alpha, \omega)' \in \mathbb{R}^2 \mid 0 \leq \alpha \leq \alpha_U, 0 < \omega_L \leq \omega \leq \omega_U\}$  (cf. Andrews, 2001, eq. (2.2)), which restricts the ARCH parameter  $\alpha$  to be nonnegative and bounds the conditional variance intercept parameter,  $\omega$ , away from zero. Note that this is a special case of the GARCH example in Andrews (2001), and as there the true parameter  $\theta_0 := (\alpha_0, \omega_0)'$  is assumed to be in  $\Theta$  such that the restrictions defining  $\Theta$

are only binding possibly at  $\alpha_0 = 0$  (this excludes, without loss of generality, cases where  $\omega_0 = \omega_L$  or  $\omega_0 = \omega_U$ ). Denote by  $\hat{\theta}_T = (\hat{\alpha}_T, \hat{\omega}_T)'$  the *unrestricted* quasi MLE (QMLE) maximizing  $L_T(\theta)$  over  $\Theta$ , and accordingly, by  $\tilde{\theta}_T = (\tilde{\alpha}_T, \tilde{\omega}_T)' = (0, \tilde{\omega}_T)'$  the *restricted* QMLE maximizing over  $\Theta_H \subset \Theta$ , where  $\Theta_H := \{(\alpha, \omega)' \in \mathbb{R}^2 \mid \alpha = 0, 0 < \omega_L \leq \omega \leq \omega_U\}$ .

It follows from Andrews (2001, p.711) that under standard regularity conditions, see Assumption A below, when  $H_0$  holds the quasi LR (QLR) statistic  $Q_T := 2(L_T(\hat{\theta}_T) - L_T(\tilde{\theta}_T))$  satisfies, as  $T \rightarrow \infty$ ,

$$Q_T \xrightarrow{w} Q_\infty = c(\max\{0, Z\})^2, \quad (7)$$

with  $Z$  denoting a  $N(0, 1)$  random variable and  $c := (\kappa - 1)/2$ , with  $\kappa := E(z_t^4)$ .

As for the previous i.i.d. location model case we consider the restricted bootstrap as opposed to the inconsistent unrestricted bootstrap in Andrews (2000). For the restricted bootstrap, define the bootstrap process  $\{X_t^* : t \leq T\}$ , in terms of the restricted QMLE  $\tilde{\theta}_T$ ,

$$X_t^* := \sigma_t^*(\tilde{\theta}_T)z_t^*, \quad \sigma_t^{*2}(\tilde{\theta}_T) = \tilde{\omega}_T, \quad (8)$$

with  $z_t^*$  sampled with replacement from the standardized residuals from restricted estimation. That is, with  $\tilde{z}_t := X_t/\sigma_t(\tilde{\theta}_T)$  and  $\tilde{z}_t^s := (\tilde{z}_t - \bar{\tilde{z}}_T)/(T^{-1} \sum_{t=1}^T (\tilde{z}_t - \bar{\tilde{z}}_T)^2)^{1/2}$ ,  $\bar{\tilde{z}}_T := T^{-1} \sum_{t=1}^T \tilde{z}_t$ , the bootstrap shocks  $\{z_t^* : t \leq T\}$  are drawn (with replacement) from  $\tilde{F}_T(x) := T^{-1} \sum_{t=1}^T \mathbb{I}(\tilde{z}_t^s \leq x)$  such that, conditionally on the original data,  $E^*(z_t^*) = 0$  and  $V^*(z_t^*) = 1$ . The bootstrap QLR statistic  $Q_T^*$  is

$$Q_T^* := 2(L_T^*(\hat{\theta}_T^*) - L_T^*(\tilde{\theta}_T^*)), \quad (9)$$

with  $L_T^*(\theta)$  the log-likelihood in (6) with  $X_t^*$  inserted. Moreover,  $\tilde{\theta}_T^* = (\tilde{\alpha}_T^*, \tilde{\omega}_T^*)'$  and  $\hat{\theta}_T^* = (\hat{\alpha}_T^*, \hat{\omega}_T^*)'$  maximize  $L_T^*(\theta)$  over  $\Theta_H$  and  $\Theta$  respectively.

In order to state the main result on consistency of the restricted bootstrap, we make the following assumptions.

#### ASSUMPTION A:

- A.1.  $\{X_t : t \geq 1\}$  is stationary and ergodic, and  $\{z_t : t \geq 1\}$  is i.i.d. with  $E(z_t) = 0$  and  $V(z_t) = 1$ .
- A.2.  $\{z_t : t \geq 1\}$  has finite fourth order moments; that is,  $\kappa := E(z_t^4) < \infty$ .

Under the null hypothesis,  $X_t = \omega^{1/2}z_t$  and Assumption A implies that also  $\{X_t : t \geq 1\}$  has finite fourth order moments. In order to investigate the properties of the bootstrap under the alternative as well, we require that finiteness of the fourth order moments holds under the alternative as well. This is done in Assumption B below.

#### ASSUMPTION B:

$\{X_t : t \geq 1\}$  has finite fourth order moments such that  $\kappa^\dagger := E(X_t^4)/(E(X_t^2))^2 < \infty$ .



Assumption A is needed for asymptotic inference when a parameter is on the boundary of the parameter space (see Andrews, 2001), and provides sufficient conditions for establishing the asymptotic properties of the QMLEs  $\hat{\theta}_T$ ,  $\tilde{\theta}_T$  and the QLR statistic  $Q_T$  when  $H_0$  holds. Assumption B differs from classic ARCH QMLE theory which excludes parameters on the boundary, see for example Jensen and Rahbek (2004) and Kristensen and Rahbek (2005). This is for two reasons: (i) we consider hypothesis testing with a parameter on the boundary, and (ii) we establish validity of the bootstrap both under the null *and* the alternative. Specifically, Assumption B imposes finite fourth order moments in order to establish the asymptotic properties of the bootstrap estimators  $\hat{\theta}_T^*$  and  $\tilde{\theta}_T^*$ , and of the bootstrap test-statistic  $Q_T^*$  when  $H_0$  does not hold, see also Remarks 3.2 and 3.3.

Theorem 2 states that  $Q_T^*$  converges weakly (in probability) to  $Q_\infty$  when the true parameter  $\theta_0$  satisfies the null  $H_0$ , while it preserves the key property of being of order  $O_p^*(1)$ , in probability, under the alternative hypothesis. That is, also the restricted bootstrap for the ARCH case is consistent.

**THEOREM 2** *Consider the bootstrap QLR statistic  $Q_T^*$  defined in (9). Under Assumption A, with the true parameter  $\theta_0 \in \Theta_H$ ,  $Q_T^* \xrightarrow{w^*}_p Q_\infty$ , where  $Q_\infty$  is given by (7). Under Assumptions A and B, with the true parameter  $\theta_0 \in \Theta \setminus \Theta_H$ ,  $Q_T^* \xrightarrow{w^*}_p \varrho Q_\infty$  where  $\varrho := (\kappa^\dagger - 1)/(\kappa - 1) < \infty$ , such that  $Q_T^* = O_p^*(1)$ , in probability.*

The proof of Theorem 2 in Appendix A is based on modifying the proof in Andrews (2001, Appendix D and Section 5.2) for the bootstrap. Thus we first verify consistency of the bootstrap (un-)restricted QML estimators, and next derive the limiting behavior of the first and second order derivatives of the bootstrap log-likelihood function, in addition to verifying details of the parameter space allowing the non-standard expansion of first derivative of the log-likelihood function at the boundary. A key difference from the non-bootstrap case is that in order to derive the bootstrap validity, we analyze the behavior of the restricted and unrestricted bootstrap QML estimators both under the hypothesis and the alternative. Also, as the bootstrap sample is generated with the restricted estimator  $\tilde{\theta}_T$  as bootstrap true parameter value, the log-likelihood function is expanded around the restricted estimator, while (weakly in probability) limit distributions are characterized in terms of a pseudo-true parameter  $\theta_0^\dagger$ , defined in Appendix A, eq. (A.1).

We illustrate the results of Theorem 2 by simulations detailed in Section 4 which show that indeed the proposed bootstrap test has correct empirical size for even small samples, while the asymptotic test is undersized, or conservative. Moreover, in terms of power, there is no discernible difference between the bootstrap and the (size-corrected) asymptotic tests.

**REMARK 3.1** The computation of the bootstrap p-value, say  $p_T^*$ , associated with (9), requires generating  $B$  (conditionally) independent bootstrap statistics,  $Q_{T,b}^*$ ,  $b = 1, \dots, B$ ;  $p_T^*$  is then approximated by  $\tilde{p}_{T,B}^* := B^{-1} \sum_{b=1}^B \mathbb{I}(Q_{T,b}^* > Q_T)$ , and is such that  $\tilde{p}_{T,B}^* \rightarrow p_T^*$  (almost surely) as  $B \rightarrow \infty$ ; cf. Hansen (1996), Andrews and Buchinsky (2000) and Davidson and MacKinnon (2000).

REMARK 3.2 As to Assumption B, note that asymptotic theory for the restricted QML estimator  $\tilde{\theta}_T$  under the alternative requires second order moments for consistency to the pseudo-true value  $\theta_0^\dagger$ , see Lemma A.1 in the appendix, while finite fourth order moments as in Assumption B are needed to establish asymptotic normality.

REMARK 3.3 One may relax Assumption B to finite second order moments, if the bootstrap algorithm is modified by combining the restricted parameter estimates with *unrestricted* residuals. More specifically, suppose that the bootstrap process as before is defined in terms of the *restricted* QMLE  $\tilde{\theta}_T$  as in (8), but with the bootstrap innovations  $z_t^*$  sampled from standardized residuals, say  $\hat{z}_t^s$ , obtained from *unrestricted* estimation: that is,  $\hat{z}_t^s$  are defined as  $\tilde{z}_t^s$  but with  $\tilde{z}_t$  replaced by  $\hat{z}_t$ , where  $\hat{z}_t := X_t/\sigma_t(\hat{\theta}_T)$ . In this case, it follows that under Assumption A and the assumption of finite second order moments of the ARCH process, consistency as in Theorem 2 holds with  $\varrho = 1$ . This follows by identical arguments used to establish Theorem 2 by replacing Lemma B.1 with Lemma B.2 in the appendix. Importantly, note that compared to the simulations reported in Appendix 4 for restricted residuals, (unreported) simulations with unrestricted residuals imply a marginally larger empirical size.

## 4 MONTE CARLO SIMULATIONS

In this section small-sample properties of the proposed bootstrap test are investigated by Monte Carlo simulations for the ARCH process in (5) with different parameter values under the null and under the alternative. Overall, the simulations show excellent size and power properties of the suggested bootstrap test based on restricted estimates and residuals, even for very small samples. Also, the bootstrap test is superior to (i) the asymptotic test based on Andrews (2001), and (ii) to the familiar Lagrange multiplier (LM) test for homoskedasticity (Engle, 1982), even when a bootstrap version of the LM test is considered. Note that based on simulations not included, the use of unrestricted residuals for the bootstrap algorithm, see Remark 3.3, imply that size is marginally increased when compared to the proposed use of restricted residuals. Apart from this, the conclusions remain identical whether or not restricted or unrestricted residuals are used.

The Monte Carlo experiment is based on parameter values  $\theta_0 = (\alpha_0, \omega_0)'$  both under the null and under the alternative hypotheses. The ARCH process is initiated at  $X_0 = 0$  and its innovations are either Gaussian or  $t$ -distributed (with  $\nu = 5$  degrees of freedom). Samples of size  $T \in \{50, 100, 500, 1000\}$  and nominal significance levels of 2.5%, 5%, and 10% are considered. The number of Monte Carlo replications is 10,000, while  $B = 399$  bootstrap repetitions are used for approximating the distribution of  $Q_T^*$  in (9), see Remark 3.1.<sup>1</sup> When comparing the bootstrap test with the asymptotic test it is important to account for the fact that, under  $H_0$ ,  $Q_T$  converges weakly to  $Q_\infty = c(\max\{0, Z\})^2$ , with  $Z \stackrel{d}{=} N(0, 1)$ ,  $c = (\kappa - 1)/2$  and  $\kappa := Ez_t^4$ ; see (7). Therefore,

---

<sup>1</sup>All computations have been performed using Ox 7.10, see Doornik (2007). Code is available upon request.

following Andrews (2001) the asymptotic test is based on the rescaled statistic

$$\tilde{Q}_T := \tilde{c}_T^{-1} Q_T \xrightarrow{w} (\max\{0, Z\})^2,$$

where  $\tilde{c}_T := \frac{1}{2}(T^{-1} \sum_{t=1}^T \tilde{z}_t^{s4} - 1)$  is a consistent estimator of  $c$  under the null hypothesis (Andrews, 2001, eq. (5.16)). The Monte Carlo results do not qualitatively differ if  $\tilde{c}_T$  is fixed at (the empirically infeasible value)  $\tilde{c}_T = c$ .

We also compare the bootstrap and asymptotic version of the QLR test with the LM test of Engle (1982), based on the auxiliary regression,

$$(X_t - \bar{X}_T)^2 = \delta_0 + \delta_1(X_{t-1} - \bar{X}_T)^2 + \eta_t, \quad t = 1, \dots, T, \quad (10)$$

with  $\bar{X}_T := (T+1)^{-1} \sum_{t=0}^T X_t$ . The test statistic for homoskedasticity is  $TR^2$ , where  $R^2$  is the coefficient of determination from (10), which is asymptotically distributed as a  $\chi^2(1)$  under the null, see Engle (1982). In addition to the asymptotic  $\chi^2(1)$  test, we also consider a bootstrap version of the LM test, see Gel and Chen (2012).

**EMPIRICAL REJECTION FREQUENCIES UNDER THE NULL.** Empirical rejection frequencies (ERFs) under the null hypothesis are reported in panel (a) of Tables I and II for Gaussian and  $t$ -distributed innovations respectively and ARCH parameter values  $\theta_0 = (\alpha_0, \omega_0)' = (0, 1)'$ .

For the Gaussian case in Table I, the bootstrap test has excellent size properties, with ERFs very close the nominal levels. This is the case even for sample sizes as small as  $T = 50$ . On the contrary, the asymptotic QLR test and asymptotic LM tests are conservative in small samples, with ERFs much below the corresponding nominal levels. The bootstrap version of the LM test has fine size properties, although it is slightly conservative.

For  $t$ -distributed innovations, Table II, the performance of the bootstrap is again excellent, with only a minor size deterioration in very small samples with respect to the Gaussian case.

**EMPIRICAL REJECTION FREQUENCIES UNDER THE ALTERNATIVE.** ERFs under the alternative are reported in Tables I and II, for Gaussian and  $t$ -distributed errors, respectively. We consider a sequence of points  $\theta_0$  under the alternative by choosing  $\alpha_0$  from the set  $\{0.1, \dots, 0.7\}$  with  $\omega_0 = 1$ . For each parameter point  $\theta_0 = (\alpha_0, \omega_0)'$  we report pointwise size adjusted ERFs in panel (c) and non-adjusted ERFs in panel (b). Also, in order to challenge the robustness of our statistical analysis and conclusions, parameter configurations with  $\alpha_0$  large,  $\alpha_0 \geq 1/\sqrt{Ez_t^4}$ , imply that Assumption B is violated.

The pointwise size adjustments<sup>2</sup> imply that the ERFs under the alternative for the different tests are directly comparable as they have identical empirical size, see also Davidson and MacKinnon (2006). As reflected in Tables I and II the asymptotic versions

---

<sup>2</sup>The size adjustments are constructed by choosing a nominal level that for given sample length  $T$  would have given the desired, say 10% rejection frequency under the null hypothesis. This adjusted nominal level is then used for parameters under the alternative. The size-adjustment is by construction infeasible in practice, but as noted it allows direct comparisons of the power properties.

of the QLR test and the LM test are quite conservative for finite samples, and hence the size adjusted rejection frequencies reported in panel (c) differ from the unadjusted ERFs under the alternative. Similarly, size adjusted and unadjusted bootstrap tests ERFs are almost identical due to the excellent size properties implying minor, if any, size adjustments.

In terms of power, or ERFs under the alternative, we note that there is no discernible difference between the bootstrap test and the corresponding (size-adjusted) asymptotic test, and moreover, the QLR tests have markedly higher power than the (bootstrap) LM test. Also the imposed violation of Assumption B seems not to be crucial.

REMARK 4.1 Unreported simulations show that, in line with the unrestricted bootstrap for the location model, see (2), that a bootstrap test based on the unrestricted ARCH parameters as expected performs poorly, in particular also when compared with the restricted bootstrap reported here.

## 5 CONCLUSION

In this paper we have established – for the leading examples discussed in Andrews (2000) – the result that consistency of bootstrap hypothesis testing when a parameter may lie on the boundary of a parameter space can be obtained by simply defining the bootstrap data generating process in terms of parameter estimators restricted by the null hypothesis. The corresponding ‘restricted bootstrap’, which as mentioned has recently been applied to solve inference problems in non-standard settings such as hypothesis testing in nonstationary co-integrated models (Cavaliere, Rahbek and Taylor, 2012; Cavaliere, Nielsen and Rahbek, 2015; Boswijk, Cavaliere, Rahbek and Taylor, 2016), is therefore argued to be an important tool for approximating limiting distributions of parameter estimators and test statistics in the case of inference in models involving parameters on the boundary of the parameter set.

Consistency of the restricted bootstrap may be extended to more involved and empirically relevant testing problems in which the true parameter is on the boundary of the parameter space. However, in general when considering likelihood-based inference and the bootstrap, we emphasize, in line with Andrews (2000), that the restricted bootstrap is indeed not a universal tool in non-standard settings, but should be studied carefully for each case by verifying consistency of the bootstrap.

We conclude by stressing that the focus of the present paper is on bootstrap hypothesis testing based on parameter estimates restricted by the null hypothesis. A natural continuation of our work could be to investigate whether the restricted bootstrap may be a helpful device for constructing (uniformly valid) confidence sets for a parameter which is not necessarily in the interior of the parameter space. In this framework, one could in principle consider confidence sets based on (restricted bootstrap) test inversion, as done e.g. in Hansen (1999), where a variant of the restricted bootstrap, the so-called grid bootstrap, was successfully used to deliver confidence sets for the largest autoregressive root in AR processes with a possible unit root. This is left out for future research.

TABLE I  
EMPIRICAL REJECTION FREQUENCIES FOR THE BOOTSTRAP AND THE  
ASYMPTOTIC TESTS OF NO ARCH EFFECTS: GAUSSIAN DISTRIBUTION

(a) Size												
$T \setminus \delta$	QLR: Bootstrap			QLR: Asymptotic			LM: Bootstrap			LM: Asymptotic		
	2.5	5.0	10.0	2.5	5.0	10.0	2.5	5.0	10.0	2.5	5.0	10.0
50	2.4	5.3	10.6	1.4	3.0	7.1	2.2	4.4	9.4	1.3	2.9	6.7
100	2.7	5.4	10.6	1.7	3.6	7.7	2.3	4.8	9.7	1.7	3.6	7.8
500	2.7	5.1	10.1	2.1	4.3	8.7	2.6	5.0	9.8	2.4	4.7	9.1
1000	2.5	5.1	10.3	2.2	4.5	9.1	2.7	5.2	10.1	2.4	5.0	9.6

  

(b) Power												
$\alpha_0 \setminus T$	QLR: Bootstrap			QLR: Asymptotic			LM: Bootstrap			LM: Asymptotic		
	100	500	1000	100	500	1000	100	500	1000	100	500	1000
0.1	34.7	74.1	92.1	29.2	71.8	91.3	24.5	62.5	85.2	22.2	61.6	85.1
0.2	60.0	97.9	100.0	54.3	97.5	100.0	44.9	94.7	99.8	42.0	94.3	99.8
0.3	78.1	100.0	100.0	73.6	99.9	100.0	63.0	99.6	100.0	59.5	99.6	100.0
0.4	88.4	100.0	100.0	85.8	100.0	100.0	76.2	100.0	100.0	73.0	100.0	100.0
0.5	94.0	100.0	100.0	92.5	100.0	100.0	84.4	100.0	100.0	81.6	100.0	100.0
0.6	97.0	100.0	100.0	96.1	100.0	100.0	89.6	100.0	100.0	87.1	100.0	100.0
0.7	98.5	100.0	100.0	98.0	100.0	100.0	93.4	100.0	100.0	90.7	100.0	100.0

  

(c) Size-corrected power												
$\alpha_0 \setminus T$	QLR: Bootstrap			QLR: Asymptotic			LM: Bootstrap			LM: Asymptotic		
	100	500	1000	100	500	1000	100	500	1000	100	500	1000
0.1	33.7	74.1	92.1	33.5	74.0	92.0	24.7	62.9	85.2	24.9	62.8	85.4
0.2	59.0	97.9	100.0	59.0	97.8	100.0	45.2	94.7	99.8	45.2	94.6	99.8
0.3	77.5	100.0	100.0	77.3	100.0	100.0	63.4	99.6	100.0	63.2	99.6	100.0
0.4	88.0	100.0	100.0	87.9	100.0	100.0	76.4	100.0	100.0	75.5	100.0	100.0
0.5	93.8	100.0	100.0	93.9	100.0	100.0	84.7	100.0	100.0	83.7	100.0	100.0
0.6	96.7	100.0	100.0	96.9	100.0	100.0	89.8	100.0	100.0	88.7	100.0	100.0
0.7	98.4	100.0	100.0	98.3	100.0	100.0	93.5	100.0	100.0	92.3	100.0	100.0

Notes: Panel (a): Size properties for nominal levels 2.5%, 5%, 10% and sample length  $T \in \{50, 100, 500, 1000\}$ . Panel (b)-(c): Power properties for nominal level 10%, sample length  $T \in \{100, 500, 1000\}$  and ARCH parameter  $\alpha \in \{0.1, \dots, 0.7\}$ . The asymptotic test is based on the rescaled statistic. Empirical rejection frequencies are presented in panel (b), while pointwise size-corrected results are presented in panel (c). Based on  $B = 399$  bootstrap repetitions and 10,000 Monte Carlo replications.

TABLE II  
EMPIRICAL REJECTION FREQUENCIES FOR THE BOOTSTRAP AND THE  
ASYMPTOTIC TESTS OF NO ARCH EFFECTS:  $t$ -DISTRIBUTION

(a) Size												
$T \setminus \delta$	QLR: Bootstrap			QLR: Asymptotic			LM: Bootstrap			LM: Asymptotic		
	2.5	5.0	10.0	2.5	5.0	10.0	2.5	5.0	10.0	2.5	5.0	10.0
50	3.1	6.3	12.0	1.4	3.4	7.7	2.3	4.4	8.2	1.5	2.8	5.3
100	2.9	5.7	11.2	1.5	3.6	7.6	2.3	4.4	8.4	2.0	3.2	5.5
500	2.5	5.6	10.7	2.1	4.2	7.8	2.5	4.6	8.8	2.4	3.7	6.2
1000	2.7	5.4	10.3	2.4	4.5	8.0	2.4	4.5	8.6	2.4	3.7	6.2

  

(b) Power												
$\alpha_0 \setminus T$	QLR: Bootstrap			QLR: Asymptotic			LM: Bootstrap			LM: Asymptotic		
	100	500	1000	100	500	1000	100	500	1000	100	500	1000
0.1	30.6	61.7	81.2	23.7	55.0	76.2	22.4	52.2	72.7	17.5	46.1	66.8
0.2	49.5	90.2	98.8	41.9	87.0	98.0	38.0	83.3	96.5	31.9	78.5	94.5
0.3	64.4	98.0	99.8	57.8	97.0	99.7	51.3	94.8	99.3	45.0	91.8	98.6
0.4	75.1	99.5	100.0	70.0	99.2	99.9	62.1	98.2	99.7	55.6	96.6	99.3
0.5	83.8	99.9	100.0	79.2	99.8	100.0	70.4	99.1	99.8	63.9	98.1	99.5
0.6	89.1	100.0	100.0	85.9	100.0	100.0	77.0	99.5	99.9	70.4	98.8	99.6
0.7	92.7	100.0	100.0	90.3	100.0	100.0	81.9	99.7	100.0	75.5	99.0	99.7

  

(c) Size-corrected power												
$\alpha_0 \setminus T$	QLR: Bootstrap			QLR: Asymptotic			LM: Bootstrap			LM: Asymptotic		
	100	500	1000	100	500	1000	100	500	1000	100	500	1000
0.1	28.8	60.5	80.6	28.7	60.3	80.5	24.7	54.1	74.5	23.1	52.9	73.5
0.2	47.7	89.7	98.7	47.3	89.6	98.6	40.5	84.5	97.0	38.6	83.3	96.1
0.3	62.4	97.9	99.8	62.7	97.7	99.8	53.8	95.2	99.4	51.8	94.1	99.0
0.4	73.7	99.5	100.0	73.9	99.4	100.0	64.4	98.4	99.7	62.2	97.7	99.5
0.5	82.6	99.9	100.0	82.7	99.9	100.0	72.5	99.3	99.8	70.1	98.7	99.7
0.6	88.3	100.0	100.0	88.2	100.0	100.0	79.0	99.6	99.9	76.5	99.1	99.7
0.7	92.1	100.0	100.0	92.0	100.0	100.0	83.7	99.8	100.0	80.9	99.3	99.8

Notes: The  $t$ -distribution for the innovations has  $\nu = 5$  degrees of freedom. See also notes to Table I.

## 6 REFERENCES

- ANDREWS, D.W.K. (1999): “Estimation when a Parameter is on a Boundary”, *Econometrica*, 67(6), 1341–1383.
- (2000): “Inconsistency of the Bootstrap when a Parameter is on the Boundary of the Parameter Space”, *Econometrica*, 68(2), 399–405.
- (2001): “Testing when a Parameter is on the Boundary of the Maintained Hypothesis”, *Econometrica*, 69(3), 683–734.
- ANDREWS, D.W.K. AND M. BUCHINSKY (2000): “A Three-Step Method for Choosing the Number of Bootstrap Repetitions”, *Econometrica*, 68(1), 23–51.
- BOSWIJK, H.P., G. CAVALIERE, A. RAHBK AND A.M.R. TAYLOR (2015): “Inference on Co-integration Parameters in Heteroskedastic Vector Autoregressions”, *Journal of Econometrics*, 192, 64–85.
- CARRASCO, M. AND S. GREGOIR (2002): “Policy Evaluation in Macroeconometric Doubly Stochastic Models”, *Annales d’Économie et de Statistique*, 67/68, 73–109.
- CAVALIERE, G., H.B. NIELSEN AND A. RAHBK (2015): “Bootstrap Testing of Hypotheses on Co-Integration Relations in Vector Autoregressive Models”, *Econometrica*, 83(2), 813–831.
- CAVALIERE, G., A. RAHBK AND A.M.R. TAYLOR (2012): “Bootstrap Determination of the Co-Integration Rank in Vector Autoregressive Models”, *Econometrica*, 80(4), 1721–1740.
- DAVIDSON, R. AND J. MACKINNON (2000): “Bootstrap Tests: How Many Bootstraps?”, *Econometric Reviews*, 19(1), 55–68.
- (2006): “The Power of Bootstrap and Asymptotic Tests”, *Journal of Econometrics*, 133, 421–441.
- DOORNIK, J.A. (2007): “Object-Oriented Matrix Programming Using Ox”, 3rd ed. Timberlake Consultants Press, London.
- DVORETZKY, A. (1972): “Asymptotic Normality for Sums of Dependent Random Variables”. In L.M. Le Cam, J. Neyman and E.L. Scott (Eds.), *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Volume 2: Probability Theory, 513–535, University of California Press, Berkeley.
- ENGLE, R. F. (1982): “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation”, *Econometrica*, 50(4), 987–1008.
- GEL Y.R. AND B. CHEN (2012): “Robust Lagrange Multiplier Test for Detecting ARCH/GARCH Effects Using Permutation and the Bootstrap”, *Canadian Journal of Statistics*, 40(3), 405–426.

- GOURIÉROUX, C., A. HOLLY AND A. MONFORT (1982): “Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters”, *Econometrica*, 50(1), 63–80.
- HANSEN, B.E. (1996): “Inference When a Nuisance Parameter Is Not Identified Under the Null Hypothesis”, *Econometrica*, 64(2), 413–430.
- (1999): “The grid bootstrap and the autoregressive model”, *Review of Economics and Statistics*, 81, 594–607.
- HECKMAN, J.J., J. SMITH AND N. CLEMENTS (1997): “Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts”, *Review of Economic Studies*, 64(4), 487–535.
- HOROWITZ, J.L. (2001): “The Bootstrap in Econometrics,” in *Handbook of Econometrics*, Vol. 5, J.J. Heckman and E.E. Leamer, eds., Elsevier Science B.V., 2001, Ch. 52, pp. 3159–3228.
- JENSEN, S.T. AND A. RAHBEEK (2004): “Asymptotic Normality of the QMLE Estimator of ARCH in the Nonstationary Case”, *Econometrica*, 72(2), 641–646.
- KRISTENSEN, D. AND A. RAHBEEK (2005): “Asymptotics of the QMLE for a Class of ARCH( $q$ ) Models”, *Econometric Theory*, 21(5), 946–961.
- LANGE, T., A. RAHBEEK AND S.T. JENSEN (2011): “Estimation and Asymptotic Inference in the AR-ARCH Model”, *Econometric Reviews*, 30(2), 129–153.
- VAN DER VAART, A.W. (1998): “*Asymptotic Statistics*”, Cambridge University Press, New York.



## APPENDIX

### A PROOF OF THEOREM 2

The proof of Theorem 2 is given by verifying and modifying the arguments used to establish Andrews (2001, eq. (5.15)) for the bootstrap case. Specifically, we establish: (i) that the structure of the parameter space allows the non-standard Taylor-type expansion at the boundary; (ii) consistency of the bootstrap QML estimators; (iii) convergence of the score and information (uniformly). More precisely, (i)-(iii) are verified as follows:

(i): *On the parameter spaces  $\Theta$  and  $\Theta_H$ :*

Recall that the parameter spaces are defined as  $\Theta := \{(\alpha, \omega)' | 0 \leq \alpha \leq \alpha_U, 0 < \omega_L \leq \omega \leq \omega_U\}$  and  $\Theta_H := \{(\alpha, \omega)' | \alpha = 0, 0 < \omega_L \leq \omega \leq \omega_U\}$ . Also, recall that the true parameter  $\theta_0 := (\alpha_0, \omega_0)'$  is assumed to be in  $\Theta$  such that the restrictions defining  $\Theta$  are only binding possibly at  $\alpha_0 = 0$  (see Section 3). Moreover, as it will be clarified in the following, in order to analyze the asymptotic properties of the restricted QMLE under the alternative, it is assumed that also the pseudo-true value  $\theta_0^\dagger$  defined by

$$\theta_0^\dagger := (0, \omega_0^\dagger)', \quad (\text{A.1})$$

with  $\omega_0^\dagger := \omega_0 / (1 - \alpha_0)$ , is in  $\Theta$  and such that  $\omega_0^\dagger < \omega_U$ , which is without loss of generality as  $\omega_U$  can be chosen arbitrarily large.

Finally, with respect to the regularity conditions on the parameter spaces in Andrews (2001), the shifted parameter spaces  $\Theta - \theta_0$  and  $\Theta_H - \theta_0$  are nested in the cones  $\Lambda$  and  $\Lambda_H$  respectively, where these are defined by  $\Lambda := [0, \infty) \times \mathbb{R}$  and  $\Lambda_H := \{0\} \times \mathbb{R}$ , see also Andrews (2001, p.701 and p.711). Observe that the same holds for the sets shifted by the bootstrap pseudo-true value; that is,  $\Theta - \theta_0^\dagger$  and  $\Theta_H - \theta_0^\dagger$  are nested in  $\Lambda$  and  $\Lambda_H$  respectively. Hence, as  $\tilde{\theta}_T - \theta_0^\dagger = o_p(1)$ , which is discussed below, then for  $T$  large enough, this holds for  $\tilde{\theta}_T$  as well such that the likelihood expansion in Andrews (2001) can be applied around the bootstrap true-value  $\tilde{\theta}_T$  instead of the (sample) true parameter  $\theta_0$ .

(ii): *On consistency of the bootstrap QML estimators:*

A key assumption in Andrews (2001, Appendix D, verification of Assumption 1\*) is consistency of the (non-bootstrap, or standard) QML estimator to  $\theta_0$ . For the bootstrap analysis we as mentioned expand the bootstrap likelihood function around  $\tilde{\theta}_T$ , the bootstrap true value. Hence in Appendix A.1 it is established that the bootstrap QML estimators are consistent, that is,  $\tilde{\theta}_T^* - \tilde{\theta}_T = o_p^*(1)$  and  $\hat{\theta}_T^* - \tilde{\theta}_T = o_p^*(1)$ , in probability. We prove this in two steps in Appendix A.1. First we consider the standard QML estimators  $\hat{\theta}_T$  and  $\tilde{\theta}_T$ , both under the null and the alternative, and second the bootstrap QML estimators. Note that when considering asymptotics under the alternative we introduce the set  $\Theta_A := \Theta \setminus \Theta_H$ , and use the result that, with  $\theta_0 \in \Theta_A$ ,  $\tilde{\theta}_T - \theta_0^\dagger = o_p(1)$ .

(iii): *On the bootstrap score and information:*

In Appendix A.2, Lemma A.3, the results for the bootstrap score and bootstrap information replace their sample equivalents (Andrews, 2001, Appendix D, verification of Assumptions 2\* and 3\*). Specifically, in Andrews (2001) the score (information) is evaluated at (in a neighborhood of)  $\theta_0$ , while here the bootstrap equivalents are evaluated at (in a neighborhood of) the bootstrap true value  $\tilde{\theta}_T$ .

Summarizing, (i)-(iii) above establish the regularity conditions for the bootstrap such that the asymptotic expansion applied in Andrews (2001, eq. (5.15)) in terms of the score and information hold for the restricted bootstrap. Specifically as in Andrews (2001, eq. (5.15)) for  $\theta_0 \in \Theta$  the results on the score and information in Lemma A.3 imply that for the bootstrapped LR statistic  $Q_T^* = 2(L_T^*(\hat{\theta}_T^*) - L_T^*(\tilde{\theta}_T^*))$  in (9),

$$Q_T^* \xrightarrow{w_p} c^\dagger (\max(0, Z))^2,$$

under Assumptions A and B with  $c^\dagger = (\kappa^\dagger - 1) / 2$ . Next, use that by definition  $c^\dagger = c\varrho$  with  $\varrho = (\kappa^\dagger - 1) / (\kappa - 1)$  and hence  $Q_T^* \xrightarrow{w_p} \varrho Q_\infty$  as claimed. Finally, observe that when  $\theta_0 \in \Theta_H$ , Assumption A implies Assumptions B and B' hold (and  $\kappa^\dagger = \kappa$ , cf. Remark A.1) and  $\varrho = 1$ .

## A.1 CONSISTENCY OF (STANDARD AND) BOOTSTRAP QML ESTIMATORS

When establishing consistency we can relax Assumption B as follows (see also Remark 3.3):

ASSUMPTION B':

$\{X_t : t \geq 1\}$  has finite second order moments.

Consider first consistency of the QML estimators  $\hat{\theta}_T$  and  $\tilde{\theta}_T$ :

LEMMA A.1 *The following results hold as  $T \rightarrow \infty$ :*

- (i) *If  $\theta_0 \in \Theta$ , and under Assumption A.1, then  $\hat{\theta}_T - \theta_0 = o_p(1)$ ;*
- (ii) *If  $\theta_0 \in \Theta_H$ , and under Assumption A.1, then  $\tilde{\theta}_T - \theta_0 = o_p(1)$ ;*
- (iii) *If  $\theta_0 \in \Theta_A$  and under Assumptions A.1 and B',  $\tilde{\theta}_T - \theta_0^\dagger = o_p(1)$ , where  $\theta_0^\dagger := (0, \omega_0^\dagger)' \in \Theta_H$ ,  $\omega_0^\dagger := \omega_0 / (1 - \alpha_0)$ .*

Observe that a bootstrap sample generated with bootstrap true value  $\theta_0^\dagger$  will be (conditionally on the original data) i.i.d. with mean zero, constant variance  $\omega_0^\dagger$  and will satisfy the null hypothesis.

The next lemma extends Lemma A.1 to the case of the bootstrap QML estimators  $\hat{\theta}_T^*$  and  $\tilde{\theta}_T^*$ . In particular, it demonstrates that they converge, as  $T \rightarrow \infty$ , to the (limiting bootstrap) pseudo true value,  $\theta_0^\dagger$ .

LEMMA A.2 Under Assumptions A.1 and B', the following results hold as  $T \rightarrow \infty$ :

- (i) If  $\theta_0 \in \Theta_H$ ,  $\hat{\theta}_T^* - \theta_0 = o_p^*(1)$  and  $\tilde{\theta}_T^* - \theta_0 = o_p^*(1)$ , in probability.
- (ii) If  $\theta_0 \in \Theta_A$ ,  $\hat{\theta}_T^* - \theta_0^\dagger = o_p^*(1)$  and  $\tilde{\theta}_T^* - \theta_0^\dagger = o_p^*(1)$ , in probability.

#### A.1.1 PROOF OF LEMMA A.1

For the unrestricted estimator  $\hat{\theta}_T$  the result holds for  $\theta_0 \in \Theta$  by Kristensen and Rahbek (2005, Theorem 1). For the restricted estimator, note that  $\tilde{\theta}_T = (0, \tilde{\omega}_T)'$  with (for  $T$  large enough)  $\tilde{\omega}_T = T^{-1} \sum_{t=1}^T X_t^2$ , such that by standard application of the law of large numbers,  $\tilde{\omega}_T \xrightarrow{p} EX_t^2 < \infty$  for any  $\theta_0 \in \Theta$  by Assumption B'. In particular, with  $\theta_0 \in \Theta_H$ ,  $EX_t^2 = \omega_0$ , while if  $\theta_0 \in \Theta_A$ ,  $EX_t^2 = \omega_0/(1 - \alpha_0) = \omega_0^\dagger$ .

#### A.1.2 PROOF OF LEMMA A.2

Note initially that, for  $T$  large enough,  $\tilde{\theta}_T^* = (0, \tilde{\omega}_T^*)'$  with  $\tilde{\omega}_T^* := T^{-1} \sum_{t=1}^T X_t^{*2} = \tilde{\omega}_T + \tilde{\omega}_T(T^{-1} \sum_{t=1}^T (z_t^{*2} - 1))$ . Conditionally on the original data,  $z_t^{*2} - 1$  is i.i.d. with mean zero, such that by standard bootstrap arguments  $\tilde{\omega}_T^* = \tilde{\omega}_T + o_p^*(1)$ , in probability, see also Lemma B.3 below. The fact that under Assumption B',  $\tilde{\omega}_T = \omega_0^\dagger + o_p(1)$  finally implies  $\tilde{\omega}_T^* = \omega_0^\dagger + o_p^*(1)$ , and hence the claimed result for  $\tilde{\theta}_T^*$  holds.

For the unrestricted bootstrap estimator  $\hat{\theta}_T^*$ , we establish in probability,

$$\sup_{\theta \in \Theta} \left| \frac{1}{T} L_T^*(\theta) - E_{\theta_0^\dagger} \ell_t(\theta) \right| = o_p^*(1), \quad (\text{A.2})$$

where  $E_{\theta_0^\dagger}$  denotes the expectation under the probability measure indexed by  $\theta_0^\dagger$ , that is,  $-2E_{\theta_0^\dagger} \ell_t(\theta) = E(\log(\omega + \alpha \omega_0^\dagger z_t^2) + \frac{\omega_0^\dagger}{\omega + \alpha \omega_0^\dagger z_t^2})$ . As  $-2E_{\theta_0^\dagger} \ell_t(\theta)$  has minimum in  $\theta_0^\dagger$ , it holds that  $\hat{\theta}_T^* - \theta_0^\dagger = o_p^*(1)$ , in probability.

To establish (A.2) use initially the triangle inequality to see that,

$$\begin{aligned} \sup_{\theta \in \Theta} \left| \frac{1}{T} L_T^*(\theta) - E_{\theta_0^\dagger} \ell_t(\theta) \right| &\leq \sup_{\theta \in \Theta} \left| \frac{1}{T} L_T^*(\theta) - E^* \ell_t^*(\theta) \right| \\ &\quad + \sup_{\theta \in \Theta} \left| E^* \ell_t^*(\theta) - E_{\theta_0^\dagger} \ell_t(\theta) \right| =: \sup_{\theta \in \Theta} |G_T^*(\theta)| + \sup_{\theta \in \Theta} |G_T(\theta)|. \end{aligned} \quad (\text{A.3})$$

We apply the uniform law of large numbers (ULLN) in Lange, Rahbek and Jensen (2011, Lemma 3) to show uniform convergence of  $G_T(\cdot)$ , and the bootstrap ULLN in Lemma B.4 for  $G_T^*(\cdot)$ . Thus for each term we first establish pointwise convergence,  $G_T(\theta) \xrightarrow{p} 0$  and  $G_T^*(\theta) \xrightarrow{p^*} 0$  respectively, as  $T \rightarrow \infty$ . Next, we establish stochastic equicontinuity as detailed below for each of the two terms.

#### POINTWISE CONVERGENCE OF $G_T(\theta)$ AND $G_T^*(\theta)$

Consider first  $G_T(\theta)$ :

We establish that  $E |G_T(\theta)| = o(1)$  which implies pointwise convergence. By definition,

$$\begin{aligned} -2(E^* \ell_t^*(\theta) - E_{\theta_0^\dagger} \ell_t(\theta)) &= \frac{1}{T} \sum_{t=1}^T \left( \log(\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2}) - E \log(\omega + \alpha \omega_0^\dagger z_t^2) \right) \\ &\quad + \frac{1}{T} \sum_{t=1}^T \left( \frac{\tilde{\omega}_T}{\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2}} - E \left( \frac{\omega_0^\dagger}{\omega + \alpha \omega_0^\dagger z_t^2} \right) \right). \end{aligned} \quad (\text{A.4})$$

A first order expansion of the  $\log(\cdot)$  term in (A.4) around  $\omega_0^\dagger$  and  $z_t^2$  gives

$$\begin{aligned} &\left| \frac{1}{T} \sum_{t=1}^T \log(\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2}) - E \log(\omega + \alpha \omega_0^\dagger z_t^2) \right| \\ &\leq \left| \frac{1}{T} \sum_{t=1}^T \log(\omega + \alpha \omega_0^\dagger z_t^2) - E \log(\omega + \alpha \omega_0^\dagger z_t^2) \right| + \frac{\alpha_U \omega_0^\dagger}{\omega_L} \left| \frac{1}{T} \sum_{t=1}^T (z_t^2 - 1) \right| + \frac{1}{\omega_L} |\tilde{\omega}_T - \omega_0^\dagger| M_T, \end{aligned}$$

where, with  $\lambda_t \in [0, 1]$ ,

$$M_T := \left| \frac{1}{T} \sum_{t=1}^T (\lambda_t \tilde{z}_t^{s2} + (1 - \lambda_t) z_t^2) \right| \leq \frac{1}{T} \sum_{t=1}^T \tilde{z}_t^{s2} + \frac{1}{T} \sum_{t=1}^T z_t^2 = 1 + \frac{1}{T} \sum_{t=1}^T z_t^2 = O_p(1)$$

where we have used the fact that  $T^{-1} \sum_{t=1}^T \tilde{z}_t^{s2} = 1$ . Observe that with  $v_t(\theta) := \log(\omega + \alpha \omega_0^\dagger z_t^2) - E \log(\omega + \alpha \omega_0^\dagger z_t^2)$ , then by the weak law of large numbers,  $T^{-1} \sum_{t=1}^T v_t(\theta) = o_p(1)$ , as  $z_t$  is i.i.d.. Likewise,  $\tilde{\omega}_T - \omega_0^\dagger = o_p(1)$  under Assumption B'. Moreover,  $T^{-1} \sum_{t=1}^T (z_t^2 - 1) = o_p(1)$ , which establishes the desired for the  $\log(\cdot)$  term.

For the remaining ratio term in (A.4), observe that

$$\frac{1}{T} \sum_{t=1}^T \left( \frac{\tilde{\omega}_T}{\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2}} - E \left( \frac{\omega_0^\dagger}{\omega + \alpha \omega_0^\dagger z_t^2} \right) \right) = \frac{1}{T} \sum_{t=1}^T \frac{(\tilde{\omega}_T - \omega_0^\dagger)}{\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2}} + \frac{1}{T} \sum_{t=1}^T \left( \frac{\omega_0^\dagger}{\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2}} - E \left( \frac{\omega_0^\dagger}{\omega + \alpha \omega_0^\dagger z_t^2} \right) \right),$$

where the first term is bounded by  $|\tilde{\omega}_T - \omega_0^\dagger|/\omega_L$  and hence  $o_p(1)$ . As above, a first order expansion around  $\omega_0^\dagger$  and  $z_t^2$  gives

$$\left| \frac{1}{T} \sum_{t=1}^T \left( \frac{\omega_0^\dagger}{\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2}} - E \left( \frac{\omega_0^\dagger}{\omega + \alpha \omega_0^\dagger z_t^2} \right) \right) \right| \leq \left| \frac{1}{T} \sum_{t=1}^T \beta_t(\theta) \right| + \frac{\alpha_U \omega_0^\dagger}{\omega_L^2} \left| \frac{1}{T} \sum_{t=1}^T (z_t^2 - 1) \right| + \frac{1}{\omega_L^2} |\tilde{\omega}_T - \omega_0^\dagger|,$$

with  $\beta_t(\theta) := \frac{1}{\omega + \alpha \omega_0^\dagger z_t^2} - E \left( \frac{1}{\omega + \alpha \omega_0^\dagger z_t^2} \right)$ , and which, as before, implies the desired for the remaining term in (A.4) and hence for the second term in (A.3).

Consider next  $G_T^*(\theta)$ :

We now establish  $G_T^*(\theta) := T^{-1} L_T^*(\theta) - E^* \ell_T^*(\theta) \xrightarrow{p^*} 0$ . By definition,

$$-2\left(\frac{1}{T} L_T^*(\theta) - E^* \ell_t^*(\theta)\right) = \frac{1}{T} \sum_{t=1}^T \left( \log(\omega + \alpha \tilde{\omega}_T z_{t-1}^{*2}) - \log(\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2}) \right)$$

$$+ \frac{1}{T} \sum_{t=1}^T \left( \frac{\tilde{\omega}_T z_t^{*2}}{\omega + \alpha \tilde{\omega}_T z_{t-1}^{*2}} - \frac{\tilde{\omega}_T}{\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2}} \right) =: K_{1,T}^* + \tilde{\omega}_T K_{2,T}^*.$$

We show that  $K_{i,T}^* = o_p^*(1)$ , in probability,  $i = 1, 2$ , such that, coupled with the fact that  $\tilde{\omega}_T = O_p(1)$ , the desired result follows.

Consider  $K_{1,T}^*$  first, and observe that it can be written as  $K_{1,T}^*(\theta) := T^{-1} \sum_{t=1}^T (\kappa_{1,t}^*(\theta) - E^*(\kappa_{1,t}^*(\theta)))$ , where  $\kappa_{1,t}^*(\theta) := \log(\omega + \alpha \tilde{\omega}_T z_{t-1}^{*2})$  with  $E^*(\kappa_{1,t}^*(\theta)) = T^{-1} \sum_{t=1}^T \log(\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2})$ . Conditionally on the sample,  $\kappa_{1,t}^*(\theta) - E^*(\kappa_{1,t}^*(\theta))$  is an i.i.d. sequence, so the conditional WLLN in Lemma B.3 applies as  $E^*|\kappa_{1,t}^*(\theta)| \leq K < \infty$  for some constant  $K$ . To see this, observe

$$E^*|\kappa_{1,t}^*(\theta)| = \frac{1}{T} \sum_{t=1}^T |\log(\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2})| \leq |\log \omega_L| + (1 + \alpha_U) \omega_U =: K < \infty.$$

Consider now  $K_{2,T}^*(\theta) := T^{-1} \sum_{t=1}^T (\kappa_{2,t}^*(\theta) - E^*(\kappa_{2,t}^*(\theta)))$ , with  $\kappa_{2,t}^*(\theta) := \frac{z_t^{*2}}{\omega + \alpha \tilde{\omega}_T z_{t-1}^{*2}}$  such that  $E^*(\kappa_{2,t}^*(\theta)) = T^{-1} \sum_{t=1}^T \frac{1}{\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2}}$ . Since (conditionally on the data)  $\kappa_{2,t}^*(\theta)$  is independent of  $\kappa_{2,t'}^*(\theta)$ ,  $|t' - t| > 1$ , we split  $K_{2,T}^*(\theta)$  as  $K_{2,T}^{*(o)}(\theta) + K_{2,T}^{*(e)}(\theta)$ , where

$$K_{2,T}^{*(o)}(\theta) := T^{-1} \sum_{t=1}^T (\kappa_{2,t}^*(\theta) - E^*(\kappa_{2,t}^*(\theta))) \mathbb{I}(t \in \{1, 3, \dots\})$$

$$K_{2,T}^{*(e)}(\theta) := T^{-1} \sum_{t=1}^T (\kappa_{2,t}^*(\theta) - E^*(\kappa_{2,t}^*(\theta))) \mathbb{I}(t \in \{2, 4, \dots\}),$$

such that each of the two terms  $K_{2,T}^{*(o)}(\theta)$  and  $K_{2,T}^{*(e)}(\theta)$  averages approximately  $T/2$  (conditionally) zero mean i.i.d. random variables. Then, we can apply the conditional WLLN of Lemma B.3 to each of the two terms  $K_{2,T}^{*(o)}(\theta)$  and  $K_{2,T}^{*(e)}(\theta)$  once it is verified that  $E^*|\kappa_{2,t}^*(\theta)| = O_p(1)$ . This follows by noticing that

$$E^*|\kappa_{2,t}^*(\theta)| = E^* \left| \frac{z_t^{*2}}{\omega + \alpha \tilde{\omega}_T z_{t-1}^{*2}} \right| = E^* |z_t^{*2}| E^* \left| \frac{1}{\omega + \alpha \tilde{\omega}_T z_{t-1}^{*2}} \right| = \frac{1}{T} \sum_{t=1}^T \frac{1}{\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2}} \leq \omega_L^{-1} < \infty,$$

thereby implying that  $G_T^*(\theta) = K_{1,T}^* + \tilde{\omega}_T (K_{2,T}^{*(o)} + K_{2,T}^{*(e)}) \xrightarrow{p^*} 0$ .

STOCHASTIC EQUICONTINUITY OF  $G_T(\theta)$  AND  $G_T^*(\theta)$

Consider first  $G_T(\theta)$ :

By the ULLN in Lange, Rahbek and Jensen (2011), the stochastic equicontinuity condition for  $G_T(\theta) = \frac{1}{T} \sum_{t=1}^T g_t(\theta)$  which implies the uniform convergence is given by

$$E \sup_{\theta \in \Theta} |2g_t(\theta)| = E \sup_{\theta \in \Theta} \left| \log(\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2}) + \frac{\tilde{\omega}_T}{\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2}} - E_{\theta_0^*} \ell_t(\theta) \right| < C < \infty.$$

From the pointwise arguments on  $G_T(\theta)$ , this simplifies to

$$E \sup_{\theta \in \Theta} (|v_t(\theta)| + |\beta_t(\theta)|) < C,$$

where  $v_t(\theta) = \log(\omega + \alpha\omega_0^\dagger z_t^2) - E \log(\omega + \alpha\omega_0^\dagger z_t^2)$ , and  $\beta_t(\theta) = \frac{1}{\omega + \alpha\omega_0^\dagger z_t^2} - E(\frac{1}{\omega + \alpha\omega_0^\dagger z_t^2})$ . Note that  $|\beta_t(\theta)| \leq 2/\omega_L$ , and  $E \sup_{\theta \in \Theta} |v_t(\theta)|$  is bounded by standard arguments for ARCH models (with true parameter  $\theta_0 := \theta_0^\dagger$ ) in Kristensen and Rahbek (2005).

Consider next  $G_T^*(\theta)$ :

By Lemma B.4 the uniform convergence is implied by showing that for all  $\theta_1, \theta_2 \in \Theta$ ,

$$|G_T^*(\theta_1) - G_T^*(\theta_2)| \leq B_T^* h(\|\theta_1 - \theta_2\|),$$

where  $E^* B_T^* = O_p(1)$  and  $h(x) \rightarrow 0$  as  $x \rightarrow 0$ . To establish this, rewrite  $G_T^*(\theta)$  as  $G_T^*(\theta) = -\frac{1}{2}(N_T^*(\theta) + M_T^*(\theta))$  with

$$\begin{aligned} M_T^*(\theta) &:= \frac{1}{T} \sum_{t=1}^T \frac{\tilde{\omega}_T z_t^{*2}}{\omega + \alpha \tilde{\omega}_T z_{t-1}^{*2}} - \frac{1}{T} \sum_{t=1}^T \frac{\tilde{\omega}_T}{\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2}} =: m_{1T}^*(\theta) - m_{2T}^*(\theta), \\ N_T^*(\theta) &:= \frac{1}{T} \sum_{t=1}^T (\log(\omega + \alpha \tilde{\omega}_T z_{t-1}^{*2}) - \log(\omega + \alpha \tilde{\omega}_T \tilde{z}_t^{s2})) =: n_{1T}^*(\theta) - n_{2T}^*(\theta). \end{aligned}$$

Consider first  $m_{2T}^*(\theta)$ :

$$\begin{aligned} |m_{2T}^*(\theta_1) - m_{2T}^*(\theta_2)| &\leq \frac{\tilde{\omega}_T}{T} \sum_{t=1}^T \left| \frac{1}{\omega_1 + \alpha_1 \tilde{\omega}_T \tilde{z}_t^{s2}} - \frac{1}{\omega_2 + \alpha_2 \tilde{\omega}_T \tilde{z}_t^{s2}} \right| \\ &= \frac{\tilde{\omega}_T}{T} \sum_{t=1}^T \left| \frac{(\omega_2 - \omega_1) + (\alpha_2 - \alpha_1) \tilde{\omega}_T \tilde{z}_t^{s2}}{(\omega_2 + \alpha_2 \tilde{\omega}_T \tilde{z}_t^{s2})(\omega_1 + \alpha_1 \tilde{\omega}_T \tilde{z}_t^{s2})} \right| \\ &= \frac{\tilde{\omega}_T}{T \omega_L^2} \sum_{t=1}^T (|\omega_2 - \omega_1| + |\alpha_2 - \alpha_1| \tilde{\omega}_T \tilde{z}_t^{s2}) \\ &= \frac{\tilde{\omega}_T}{\omega_L^2} (|\omega_2 - \omega_1| + |\alpha_2 - \alpha_1| \tilde{\omega}_T \frac{1}{T} \sum_{t=1}^T \tilde{z}_t^{s2}) \\ &\leq \tilde{\omega}_T \frac{\max(1, \tilde{\omega}_T)}{\omega_L^2} (|\omega_2 - \omega_1| + |\alpha_2 - \alpha_1|) =: B_{m_2, T}^* h(\|\theta_1 - \theta_2\|), \end{aligned}$$

where  $h(x) = x$  and  $E^* B_{m_2, T}^* = B_{m_2, T}^* \leq \tilde{\omega}_T \frac{(1 + \tilde{\omega}_T)}{\omega_L^2} = O_p(1)$ .

Next, consider  $m_{1T}^*(\theta)$  where similarly,

$$\begin{aligned} |m_{1T}^*(\theta_1) - m_{1T}^*(\theta_2)| &\leq \left| \frac{\tilde{\omega}_T}{\omega_L^2} \frac{1}{T} \sum_{t=1}^T ((\omega_1 - \omega_2) z_t^{*2} + (\alpha_1 - \alpha_2) \tilde{\omega}_T z_{t-1}^{*2} z_t^{*2}) \right| \\ &\leq \tilde{\omega}_T \frac{\max(1, \tilde{\omega}_T)}{\omega_L^2} \max(\frac{1}{T} \sum_{t=1}^T z_t^{*2}, \frac{1}{T} \sum_{t=1}^T z_{t-1}^{*2} z_t^{*2}) (|\omega_1 - \omega_2| + |\alpha_1 - \alpha_2|) \\ &=: B_{m_1, T}^* h(\|\theta_1 - \theta_2\|), \end{aligned}$$

and  $E^* B_{m_1, T}^* = O_p(1)$  since

$$E^* \max(\frac{1}{T} \sum_{t=1}^T z_t^{*2}, \frac{1}{T} \sum_{t=1}^T z_{t-1}^{*2} z_t^{*2}) \leq \frac{1}{T} \sum_{t=1}^T E^* z_t^{*2} + \frac{1}{T} \sum_{t=1}^T E^* (z_{t-1}^{*2} z_t^{*2}) = 2.$$

Turn to  $n_{1T}^*(\theta)$ , where by the mean-value theorem with  $(a, w)' = \lambda\theta_1 + (1 - \lambda)\theta_2$  for some  $\lambda \in (0, 1)$ ,

$$\begin{aligned} |n_{1T}^*(\theta_1) - n_{1T}^*(\theta_2)| &\leq \left| \frac{1}{T} \sum_{t=1}^T \frac{(\omega_1 - \omega_2)}{(w + a\tilde{\omega}_T z_{t-1}^{*2})} + \frac{1}{T} \sum_{t=1}^T \frac{(\alpha_1 - \alpha_2)\tilde{\omega}_T z_{t-1}^{*2}}{(w + a\tilde{\omega}_T z_{t-1}^{*2})} \right| \\ &\leq \tilde{\omega}_T \frac{(1 + \tilde{\omega}_T)}{\omega_L} \left(1 + \frac{1}{T} \sum_{t=1}^T z_{t-1}^{*2}\right) (|\omega_1 - \omega_2| + |\alpha_1 - \alpha_2|) \\ &=: B_{n_1, T}^* (|\omega_1 - \omega_2| + |\alpha_1 - \alpha_2|), \end{aligned}$$

with  $E^* B_{n_1, T}^* = \tilde{\omega}_T \frac{2(1 + \tilde{\omega}_T)}{\omega_L} = O_p(1)$ . Finally, consider  $n_{2T}^*(\theta)$  where,

$$\begin{aligned} |n_{2T}^*(\theta_1) - n_{2T}^*(\theta_2)| &\leq \left| \frac{1}{T} \sum_{t=1}^T \frac{(\omega_1 - \omega_2)}{(w + a\tilde{\omega}_T \tilde{z}_t^{s2})} + \frac{1}{T} \sum_{t=1}^T \frac{(\alpha_1 - \alpha_2)\tilde{\omega}_T \tilde{z}_t^{s2}}{(w + a\tilde{\omega}_T \tilde{z}_t^{s2})} \right| \\ &\leq \frac{1 + \tilde{\omega}_T}{\omega_L} \|\theta_1 - \theta_2\| =: B_{n_2, T}^* \|\theta_1 - \theta_2\|, \end{aligned}$$

with  $E^* B_{n_2, T}^* = B_{n_2, T}^* = O_p(1)$ .

With  $B_T^* := B_{m_1, T}^* + B_{m_2, T}^* + B_{n_1, T}^* + B_{n_2, T}^*$  and  $h(x) = x$  the stochastic equicontinuity of  $G_T^*(\theta)$  thus holds as desired.

## A.2 BOOTSTRAP SCORE AND INFORMATION

Consider here the (scaled) bootstrap score and information as defined by

$$S_T^*(\theta) := \frac{1}{T} \sum_{t=1}^T s_t^*(\theta), \quad J_T^*(\theta) := \frac{1}{T} \sum_{t=1}^T j_t^*(\theta),$$

where  $s_t^*(\theta) := -2\partial\ell_t^*(\theta)/\partial\theta$  and  $j_t^*(\theta) := 2\partial^2\ell_t^*(\theta)/\partial\theta\partial\theta'$ . The following result holds irrespectively of the null hypothesis to hold.

LEMMA A.3 *Under Assumptions A and B and with  $\theta_0 \in \Theta$  it holds that as  $T \rightarrow \infty$ :*

$$T^{1/2} S_T^*(\tilde{\theta}_T) \xrightarrow{w^*}_p N(0, (\kappa^\dagger - 1)J^\dagger) \quad \text{and} \quad J_T^*(\tilde{\theta}_T) \xrightarrow{p^*}_p J^\dagger, \quad (\text{A.5})$$

where

$$J^\dagger := \frac{1}{\omega_0^{\dagger 2}} \Sigma^\dagger, \quad \Sigma^\dagger := \begin{pmatrix} 1 & \omega_0^\dagger \\ \omega_0^\dagger & \omega_0^{\dagger 2} \kappa^\dagger \end{pmatrix} \quad (\text{A.6})$$

with  $\kappa^\dagger$ , defined in Assumption B, equal to  $\kappa^\dagger = \kappa(1 - \alpha_0^2)/(1 - \kappa\alpha_0^2)$ . Moreover, the convergence of  $J_T^*(\cdot)$  holds uniformly (in probability); that is,

$$\sup_{\theta \in \Theta} |J_T^*(\theta) - J^\dagger| \xrightarrow{p^*}_p 0. \quad (\text{A.7})$$

REMARK A.1 Note that for  $\theta_0 \in \Theta_H$ , Assumption A implies Assumption B holds and the results in Lemma A.3 hold with the entries  $\omega_0^\dagger$  and  $\kappa^\dagger$  in  $\Sigma^\dagger$  reducing to  $\omega_0$  and  $\kappa$ , respectively.

PROOF OF LEMMA A.3: For the convergence to the Gaussian distribution in (A.5), notice first that by definition,  $s_t^*(\theta) = \left(1 - \frac{X_t^{*2}}{\omega + \alpha X_{t-1}^{*2}}\right) \frac{1}{\omega + \alpha X_{t-1}^{*2}} (1, X_{t-1}^{*2})'$ , such that

$$s_t^*(\tilde{\theta}_T) = (1 - z_t^{*2}) \frac{1}{\tilde{\omega}_T} (1, X_{t-1}^{*2})'.$$

With  $\mathcal{F}_{t-1}^* := \sigma(X_{t-i}^* : i \geq 1)$ , it follows that  $E^*(s_t^*(\tilde{\theta}_T) \mid \mathcal{F}_{t-1}^*) = 0$  and we verify classic regularity conditions for the CLT for martingale difference arrays in Dvoretzky (1972, Theorem 2.2) in the bootstrap case. The conditional variance is given by

$$\begin{aligned} E^*(s_t^*(\tilde{\theta}_T) s_t^*(\tilde{\theta}_T)' \mid \mathcal{F}_{t-1}^*) &= E^*((1 - z_t^{*2})^2 \frac{1}{\tilde{\omega}_T^2} (1, X_{t-1}^{*2})' (1, X_{t-1}^{*2}) \mid \mathcal{F}_{t-1}^*) \\ &= \frac{1}{\tilde{\omega}_T^2} (1, X_{t-1}^{*2})' (1, X_{t-1}^{*2}) \frac{1}{T} \sum_{t=1}^T (1 - \tilde{z}_t^{s2})^2. \end{aligned}$$

By Lemma B.1,  $T^{-1} \sum_{t=1}^T (1 - \tilde{z}_t^{s2})^2 \xrightarrow{p} \kappa^\dagger - 1$ . Moreover, by Lemma A.1,  $\tilde{\omega}_T \xrightarrow{p} \omega_0^\dagger$  and,

$$\frac{1}{T} \sum_{t=1}^T (1, X_{t-1}^{*2})' (1, X_{t-1}^{*2}) \xrightarrow{p} \Sigma^\dagger := \begin{pmatrix} 1 & \omega_0^\dagger \\ \omega_0^\dagger & \omega_0^{\dagger 2} \kappa^\dagger \end{pmatrix}$$

by Lemma B.3. Collecting terms, the average conditional variance of  $s_t^*(\tilde{\theta}_T)$  satisfies

$$\frac{1}{T} \sum_{t=1}^T E^*(s_t^*(\tilde{\theta}_T) s_t^*(\tilde{\theta}_T)' \mid \mathcal{F}_{t-1}^*) \xrightarrow{p} (\kappa^\dagger - 1) J^\dagger,$$

with  $J^\dagger = \omega_0^{\dagger -2} \Sigma^\dagger$ . Finally, the Lindeberg condition holds by showing that for any  $\epsilon > 0$  and  $\lambda = (\lambda_1, \lambda_2)' \in \mathbb{R}^2$ ,

$$\frac{1}{T} \sum_{t=1}^T E^*((\lambda' s_t^*(\tilde{\theta}_T))^2 \mathbb{I}(|\lambda' s_t^*(\tilde{\theta}_T)| > \epsilon T^{1/2})) = E^*((\lambda' s_t^*(\tilde{\theta}_T))^2 \mathbb{I}(|\lambda' s_t^*(\tilde{\theta}_T)| > \epsilon T^{1/2})) \xrightarrow{p} 0$$

where the equality holds by (conditional) stationarity. First, observe that by definition  $\lambda' s_t^*(\tilde{\theta}_T) = \tilde{\omega}_T^{-1} (1 - z_t^{*2}) (\lambda_1 + \lambda_2 \tilde{\omega}_T z_{t-1}^{*2})$ . Using the standard inequality  $|xy| \leq x^2 + y^2$ , we can use the bound  $E^*((\lambda' s_t^*(\tilde{\theta}_T))^2 \mathbb{I}(|\lambda' s_t^*(\tilde{\theta}_T)| > \epsilon T^{1/2})) \leq \xi_{1,T} + \xi_{2,T}$ , where

$$\begin{aligned} \xi_{1,T} &:= \frac{1}{\tilde{\omega}_T} E^*((\lambda' s_t^*(\tilde{\theta}_T))^2 \mathbb{I}(|(1 - z_t^{*2}) \frac{1}{\tilde{\omega}_T}|^2 > \epsilon T^{1/2})) \\ \xi_{2,T} &:= \frac{1}{\tilde{\omega}_T} E^*((\lambda' s_t^*(\tilde{\theta}_T))^2 \mathbb{I}(|\lambda_1 + \lambda_2 \tilde{\omega}_T z_{t-1}^{*2}|^2 > \epsilon T^{1/2})). \end{aligned}$$

By (conditional) independence of  $z_t^*$  and  $z_{t-1}^*$ ,

$$\begin{aligned} \xi_{1,T} &= \frac{1}{\tilde{\omega}_T} E^*((1 - z_t^{*2})^2 \mathbb{I}(|(1 - z_t^{*2}) \frac{1}{\tilde{\omega}_T}|^2 > \epsilon T^{1/2})) E^*((\lambda_1 + \lambda_2 \tilde{\omega}_T z_{t-1}^{*2})^2) \\ &= o_p(1) O_p(1) \end{aligned}$$

since  $\tilde{\omega}_T^{-1} E^*((\lambda_1 + \lambda_2 \tilde{\omega}_T z_{t-1}^{*2})^2) = \tilde{\omega}_T^{-1} T^{-1} \sum_{t=1}^T (\lambda_1 + \lambda_2 \tilde{\omega}_T \tilde{z}_t^{s2})^2 = O_p(1)$  by applying Lemma B.1(i), while

$$E^*((1 - z_t^{*2})^2 \mathbb{I}(|(1 - z_t^{*2}) \frac{1}{\tilde{\omega}_T}|^2 > \epsilon T^{1/2}))$$



$$= \frac{1}{T} \sum_{t=1}^T (1 - \tilde{z}_t^{s2})^2 \mathbb{I} \left( \left| (1 - \tilde{z}_t^{s2}) \frac{1}{\tilde{\omega}_T} \right|^2 > \epsilon T^{1/2} \right) \xrightarrow{p} 0$$

by using Lemma B.1(ii) with  $q = 1/4$ . By the same arguments we also find that  $\xi_{2,T} = o_p(1)$ . This verifies the Lindeberg condition, and therefore that  $T^{1/2} S_T^*(\tilde{\theta}_T) \xrightarrow{w^*}_p N(0, (\kappa^\dagger - 1)J^\dagger)$ .

Turning to the information, we have that, by definition,  $J_T^*(\theta) = T^{-1} \sum_{t=1}^T j_t^*(\theta)$  with

$$j_t^*(\theta) = \left( 2 \frac{X_t^{*2}}{\omega + \alpha X_{t-1}^{*2}} - 1 \right) \frac{1}{\omega + \alpha X_{t-1}^{*2}} (1, X_{t-1}^{*2}) (1, X_{t-1}^{*2})'.$$

Hence as for the score,

$$J_T^*(\tilde{\theta}_T) = \frac{1}{\tilde{\omega}_T} \frac{1}{T} \sum_{t=1}^T (2z_t^{*2} - 1) (1, X_{t-1}^{*2}) (1, X_{t-1}^{*2})' \xrightarrow{p^*}_p \frac{1}{\omega_0^{\dagger 2}} \Sigma^\dagger = J^\dagger,$$

which establishes (A.5). Next, to establish (A.7), we apply Lemma B.4. Observe first that

$$|j_t^*(\theta)| = \left| \left( 2 \frac{X_t^{*2}}{\omega + \alpha X_{t-1}^{*2}} - 1 \right) \frac{1}{\omega + \alpha X_{t-1}^{*2}} (1, X_{t-1}^{*2})' (1, X_{t-1}^{*2}) \right| \leq \delta_T^*,$$

with  $\delta_T^* := \left( \frac{2X_t^{*2}}{\omega_L} + 1 \right) \left( \frac{1+X_{t-1}^{*4}}{\omega_L} \right)$ . And moreover,

$$E^* \delta_T^* = E^* \left( \left( \frac{2\tilde{\omega}_T z_t^{*2}}{\omega_L} + 1 \right) \left( \frac{1+\tilde{\omega}_T^2 z_{t-1}^{*4}}{\omega_L} \right) \right) = \left( \frac{2\tilde{\omega}_T}{\omega_L} + 1 \right) \frac{1}{\omega_L} \left( 1 + \tilde{\omega}_T^2 \frac{1}{T} \sum_{t=1}^T \tilde{z}_t^{s4} \right) \xrightarrow{p} \delta,$$

with  $\delta := \left( \frac{2\omega_0^\dagger}{\omega_L} + 1 \right) \left( \frac{1+\omega_0^{\dagger 2} \kappa^\dagger}{\omega_L} \right)$ , which establishes (A.7).  $\square$

## B AUXILIARY LEMMAS

In this section we list some auxiliary lemmas used in the previous. These include a bootstrap weak LLN and a bootstrap uniform law.

**LEMMA B.1** *With  $\{\tilde{z}_t^s : t \geq 1\}$  defined in Section 3, under Assumptions A and B, as  $T \rightarrow \infty$ , (i)  $T^{-1} \sum_{t=1}^T \tilde{z}_t^{s4} \xrightarrow{p} \kappa^\dagger$  and (ii) for any  $q \in (0, \infty)$  and for all  $\epsilon > 0$ ,  $\tilde{N}_T := T^{-1} \sum_{t=1}^T \tilde{z}_t^{s4} \mathbb{I}(|\tilde{z}_t^s| > \epsilon T^q) \xrightarrow{p} 0$ .*

**PROOF:** For Part (i), using the equality  $\tilde{z}_t = \tilde{\omega}_T^{-1/2} X_t$ , we have that

$$T^{-1} \sum_{t=1}^T \tilde{z}_t^{s4} = \left( T^{-1} \sum_{t=1}^T (X_t - \bar{X}_T)^2 \right)^{-2} T^{-1} \sum_{t=1}^T (X_t - \bar{X}_T)^4 \xrightarrow{p} \kappa^\dagger$$

under Assumptions A and B.

For Part (ii), we first show that

$$N_T := \frac{1}{T} \sum_{t=1}^T X_t^4 \mathbb{I}(|X_t| > \epsilon T^q) \xrightarrow{p} 0 \tag{B.1}$$

and next that (B.1) implies that  $\tilde{N}_T \xrightarrow{p} 0$ . The result in (B.1) follows as in van der Vaart (1998, proof of Theorem 23.4) by noting that for any  $M \leq \epsilon T^q$ , almost surely (a.s.)

$$N_T \leq \frac{1}{T} \sum_{t=1}^T (X_t^4 \mathbb{I}(X_t^2 > M)) \xrightarrow{a.s.} \gamma_M := EX_t^4 \mathbb{I}(X_t^2 > M),$$

by stationarity and ergodicity of  $X_t$  and Assumption B. Choosing  $M$  large enough  $\gamma_M$  can be made arbitrarily small and hence, for any small  $\eta$ ,  $\limsup N_T < \eta$  a.s., such that (B.1) holds.

Next, consider  $\tilde{N}_T$ . Notice that the restricted residuals satisfy  $\tilde{z}_t^s = s_{xT}^{-1}(X_t - \bar{X}_T)$  with  $s_{xT}^2 := T^{-1} \sum_{t=1}^T (X_t - \bar{X}_T)^2 \rightarrow_p \sigma_x^2 := V(X_t^2)$  under Assumptions A and B. We may also express  $s_{xT}^2$  as  $s_{xT}^2 = \sigma_x^2 (1 + B_T)$  where  $B_T := \sigma_x^{-2}(s_{xT}^2 - \sigma_x^2) = o_p(1)$ . In particular, for any  $0 < \rho < 1$ ,  $(1 + B_T) \geq \rho$  with probability tending to 1 as  $T \rightarrow \infty$ . Hence, for  $T$  large enough,

$$\begin{aligned} \mathbb{I}(|\tilde{z}_t^s| > \epsilon T^q) &= \mathbb{I}(|X_t - \bar{X}_T| > \epsilon s_{xT}^2 T^q) = \mathbb{I}(|X_t - \bar{X}_T| > \epsilon \sigma_x^2 T^q (1 + B_T)) \\ &\leq \mathbb{I}(|X_t - \bar{X}_T| > c_\rho T^q), \end{aligned}$$

with  $c_\rho := \epsilon \sigma_x^2 \rho > 0$ . This again implies that

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T \tilde{z}_t^{s4} \mathbb{I}(|\tilde{z}_t^s| > \epsilon T^q) &\leq \frac{1}{Ts_x^2} \sum_{t=1}^T (X_t - \bar{X}_T)^4 \mathbb{I}(|X_t - \bar{X}_T| > c_\rho T^q) \\ &\leq \frac{1}{T\sigma_x^2} \sum_{t=1}^T X_t^4 \mathbb{I}(|X_t| > c_\rho T^q) + o_p(1) = \frac{1}{\sigma_x^2} N_T + o_p(1). \end{aligned}$$

To see why the last inequality holds, use that

$$\mathbb{I}(|X_t - \bar{X}_T| > c_\rho T^q) \leq \mathbb{I}(|X_t| + |\bar{X}_T| > c_\rho T^q) \leq \mathbb{I}(|X_t| > \frac{c_\rho}{2} T^q) + \mathbb{I}(|\bar{X}_T| > \frac{c_\rho}{2} T^q)$$

which implies

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X}_T)^4 \mathbb{I}(|X_t - \bar{X}_T| > c_\rho T^q) &\leq \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X}_T)^4 \mathbb{I}(|X_t| > \frac{c_\rho}{2} T^q) \\ &+ \frac{1}{T} \sum_{t=1}^T (X_t - \bar{X}_T)^4 \mathbb{I}(|\bar{X}_T| > \frac{c_\rho}{2} T^q) \end{aligned}$$

where

$$\frac{1}{T} \sum_{t=1}^T (X_t - \bar{X}_T)^4 \mathbb{I}(|\bar{X}_T| > \frac{c_\rho}{2} T^q) = \mathbb{I}(|\bar{X}_T| > \frac{c_\rho}{2} T^q) O_p(1) = o_p(1)$$

since  $\mathbb{I}(|\bar{X}_T| > \frac{c_\rho}{2} T^q) \xrightarrow{p} 0$ . Moreover, by the standard inequality  $(x + y)^4 \leq c(x^4 + y^4)$  for some finite constant  $c \geq 9$ ,

$$\frac{1}{T} \sum_{t=1}^T (X_t - \bar{X}_T)^4 \mathbb{I}(|X_t| > \frac{c_\rho}{2} T^q) \leq \frac{c}{T} \sum_{t=1}^T X_t^4 \mathbb{I}(|X_t| > \frac{c_\rho}{2} T^q) + \bar{X}_T^4 \frac{c}{T} \sum_{t=1}^T \mathbb{I}(|X_t| > \frac{c_\rho}{2} T^q)$$

where the second term on the right hand side is of  $o_p(1)$  using  $\bar{X}_T^4 = O_p(1)$  and the same arguments used for proving (B.1). This shows the missing result that  $N_T \xrightarrow{p} 0$  implies  $\tilde{N}_T \xrightarrow{p} 0$ , as required.  $\square$

LEMMA B.2 *With  $\{\hat{z}_t^s : t \geq 1\}$  defined in Section 3, Remark 3.3, under Assumptions A and B', as  $T \rightarrow \infty$ ,*

$$T^{-1} \sum_{t=1}^T \hat{z}_t^{s4} \xrightarrow{p} \kappa.$$

PROOF: By definition,  $\hat{z}_t = X_t / \sigma_t(\hat{\theta}_T) = z_t \sigma_t(\theta_0) / \sigma_t(\hat{\theta}_T)$ . Consider first the case of  $\alpha_0 > 0$ , for which it follows that with  $\delta_T = o_p(1)$ ,

$$\hat{z}_t^2 = z_t^2 (1 + \delta_T), \quad (\text{B.2})$$

such that  $T^{-1} \sum_{t=1}^T \hat{z}_t^4 \xrightarrow{p} E z_t^4 = \kappa$  under Assumption A and B'. To see (B.2), observe that,

$$z_t^2 - \hat{z}_t^2 = z_t^2 \left\{ \frac{(\hat{\omega}_T - \omega_0)}{\hat{\omega}_T + \hat{\alpha}_T X_{t-1}^2} + \frac{(\hat{\alpha}_T - \alpha_0) X_{t-1}^2}{\hat{\omega}_T + \hat{\alpha}_T X_{t-1}^2} \right\},$$

and use, with  $\delta_T := \delta_{1T} + \delta_{2T}$ ,

$$\left| \frac{\hat{\omega}_T - \omega_0}{\hat{\omega}_T + \hat{\alpha}_T X_{t-1}^2} \right| \leq \frac{|\hat{\omega}_T - \omega_0|}{|\hat{\omega}_T|} =: \delta_{1T} = o_p(1) \quad \text{and} \quad \left| \frac{(\hat{\alpha}_T - \alpha_0) X_{t-1}^2}{\hat{\omega}_T + \hat{\alpha}_T X_{t-1}^2} \right| \leq \frac{|\hat{\alpha}_T - \alpha_0|}{|\hat{\alpha}_T|} =: \delta_{2T} = o_p(1).$$

Next, consider the case when  $\alpha_0 = 0$ , where by definition

$$z_t^2 - \hat{z}_t^2 = z_t^2 \left\{ \frac{(\hat{\omega}_T - \omega_0)}{\hat{\omega}_T + \hat{\alpha}_T X_{t-1}^2} + z_{t-1}^2 \frac{\hat{\alpha}_T \omega_0}{\hat{\omega}_T + \hat{\alpha}_T X_{t-1}^2} \right\}.$$

As before,  $\left| \frac{\hat{\omega}_T - \omega_0}{\hat{\omega}_T + \hat{\alpha}_T X_{t-1}^2} \right| \leq \frac{|\hat{\omega}_T - \omega_0|}{|\hat{\omega}_T|} = \delta_{1T}$ , while  $\left| \frac{\hat{\alpha}_T \omega_0}{\hat{\omega}_T + \hat{\alpha}_T X_{t-1}^2} \right| \leq \frac{|\hat{\alpha}_T|}{|\hat{\omega}_T / \omega_0|} := \delta_{3T} = o_p(1)$ . Collecting terms,

$$\hat{z}_t^2 = z_t^2 (1 + \delta_{1T} + z_{t-1}^2 \delta_{3T}),$$

and hence

$$\frac{1}{T} \sum_{t=1}^T \hat{z}_t^4 = \frac{1}{T} \sum_{t=1}^T z_t^4 (1 + \delta_{1T} + z_{t-1}^2 \delta_{3T})^2 = \frac{1}{T} \sum_{t=1}^T z_t^4 + o_p(1),$$

where we have used that under Assumptions A and B',  $\delta_{3T}^2 T^{-1} \sum_{t=1}^T z_t^4 z_{t-1}^4 = o_p(1)$ .  $\square$

LEMMA B.3 (WLLN) *Let  $\{Z_t^* : t \leq T\}$  denote a bootstrap sample obtained by i.i.d. resampling of the triangular array  $\{Z_{T,t} : t \leq T\}$ . Suppose that  $\hat{\mu}_T := E^*(Z_t^*) \xrightarrow{p} \mu \in \mathbb{R}$  and  $E^*|Z_t^*| \leq C_T$ , where  $C_T$  is a function of  $\{Z_{T,t} : t \leq T\}$  such that  $C_T \xrightarrow{p} c \in \mathbb{R}$ . Then,  $\bar{Z}_T^* \xrightarrow{p} \mu$  and  $\bar{Z}_T^* - \hat{\mu}_T \xrightarrow{p} 0$  as  $T \rightarrow \infty$ .*

PROOF: With  $u \in \mathbb{R}$ , denote the characteristic function of  $Z_t^*$ , conditional on  $\{Z_{T,t}, t \leq T\}$ , by  $\phi^*(u) := E^*(\exp(iuZ_t^*))$ . Since  $E^*|Z_t^*| \leq C_T = O_p(1)$ , a stochastic Taylor expansion (see e.g. van der Vaart, 1998, p.15) yields

$$\phi_T^*\left(\frac{u}{T}\right) = E^*(\exp(\frac{iu}{T}Z_t^*)) = 1 + \frac{iu}{T}\hat{\mu}_T + o_p\left(\frac{1}{T}\right).$$

By the conditional independence of  $Z_t^*$ ,  $\bar{Z}_T^*$  has (conditional) characteristic function satisfying

$$\bar{\phi}_T(u) := E^*(\exp(iu\bar{Z}_T^*)) = (\phi^*(u/T))^T = (1 + \frac{iu}{T}\hat{\mu}_T + o_p(\frac{1}{T}))^T \xrightarrow{p} \exp(iu\mu).$$

Therefore,  $\bar{Z}_T^* \xrightarrow{w^*}_p \mu$ , which also implies  $\bar{Z}_T^* \xrightarrow{p^*}_p \mu$  by standard arguments. Since  $\bar{Z}_T^* - \hat{\mu}_T = \bar{Z}_T^* - \mu + (\mu - \hat{\mu}_T) = \bar{Z}_T^* - \mu + o_p(1)$ , the second statement in the lemma also holds.  $\square$

LEMMA B.4 (ULLN) *Let  $\{Z_t^* : t \leq T\}$  and  $\{Z_{T,t} : t \leq T\}$  be as in Lemma B.3. With  $G_T^*(\theta)$  a function of  $\theta \in \Theta \subseteq \mathbb{R}^k$  for some  $k \geq 1$  and of  $\{Z_t^* : t \leq T\}$ , assume that the following conditions hold:*

(i)  $\Theta$  is a compact subset of  $\mathbb{R}^k$ .

(ii) For all  $\theta \in \Theta$ ,  $G_T^*(\theta) \xrightarrow{p^*}_p 0$ .

(iii) For all  $\theta, \theta' \in \Theta$ ,

$$|G_T^*(\theta) - G_T^*(\theta')| \leq B_T^* h(\|\theta - \theta'\|) \tag{B.3}$$

where  $h(x) \rightarrow 0$  for  $x \rightarrow 0$  and  $B_T^*$  is independent of  $\theta$  and such that  $E^*(B_T^*) = O_p(1)$ .

Then, as  $T \rightarrow \infty$ ,

$$\sup_{\theta \in \Theta} |G_T^*(\theta)| \xrightarrow{p^*}_p 0.$$

PROOF: The proof follows by arguments similar to Lange, Rahbek and Jensen (2011). First notice that, since  $\Theta$  is a compact subset of  $\mathbb{R}^k$ , it holds that for any  $\delta > 0$  there exists a finite cover  $\{B(\theta_j, \delta), j = 1, \dots, J\}$  of  $\Theta$ , where  $J$  is a finite number, such that for any  $\theta_1, \theta_2 \in B(\theta_j, \delta)$ ,  $\|\theta_2 - \theta_1\| \leq \delta$  (all  $j = 1, \dots, J$ ).

By standard inequalities

$$\begin{aligned} \sup_{\theta \in \Theta} |G_T^*(\theta)| &= \sup_{\theta \in \Theta} |G_T^*(\theta) - G_T^*(\theta_j) + G_T^*(\theta_j)| \leq \sup_{\theta \in \Theta} |G_T^*(\theta) - G_T^*(\theta_j)| + |G_T^*(\theta_j)| \\ &\leq \max_{j=1, \dots, J} \sup_{\theta' \in B(\theta_j, \delta)} |G_T^*(\theta') - G_T^*(\theta_j)| + \max_{j=1, \dots, J} |G_T^*(\theta_j)|. \end{aligned}$$

Since for all  $\theta_j$ , condition (B.3) implies

$$\sup_{\theta' \in B(\theta_j, \delta)} |G_T^*(\theta') - G_T^*(\theta_j)| \leq B_T^* h(\|\theta' - \theta_j\|) \leq B_T^* h(\delta),$$

we have that

$$\sup_{\theta \in \Theta} |G_T^*(\theta)| \leq B_T^* h(\delta) + \max_{j=1, \dots, J} |G_T^*(\theta_j)|.$$

This implies that

$$P^* \left( \sup_{\theta \in \Theta} |G_T^*(\theta)| > 2\epsilon \right) \leq P^* (B_T^* h(\delta) > \epsilon) + P^* \left( \max_{j=1, \dots, J} |G_T^*(\theta_j)| > \epsilon \right), \quad (\text{B.4})$$

where by Bonferroni's inequality, the pointwise convergence assumption and the finiteness of  $J$  we have that

$$P^* \left( \max_{j=1, \dots, J} |G_T^*(\theta_j)| > \epsilon \right) \leq \sum_{j=1}^J P^* (|G_T^*(\theta_j)| > \epsilon) \xrightarrow{p} 0.$$

Regarding the first probability on the right hand side of (B.4), it suffices to notice that by the conditional Markov's inequality

$$P^* (B_T^* h(\delta) > \epsilon) \leq \frac{1}{\epsilon} E^* (B_T^* h(\delta)) = \frac{h(\delta)}{\epsilon} E^* (B_T^*),$$

and

$$P (P^* (B_T^* h(\delta) > \epsilon) > \gamma) \leq P \left( \frac{h(\delta)}{\epsilon} E^* (B_T^*) > \gamma \right) = P \left( E^* (B_T^*) \geq \frac{\epsilon \gamma}{h(\delta)} \right).$$

Since  $E^* (B_T^*)$  is an  $O_p(1)$  random variable, the latter probability can be made arbitrarily small by picking  $\delta$  small enough.  $\square$